From individual to joint species distribution models: a comparison of model complexity and predictive performance

Bogdan Caradima[*]

Nele Schuwirth

Peter Reichert

July 8, 2019

Eawag: Swiss Federal Institute of Aquatic Science and Technology
8600 Dübendorf, Switzerland
ETH Zurich, Institute of Biogeochemistry and Pollutant Dynamics
8092 Zürich, Switzerland

[*]bogdan.caradima@eawag.ch

Abstract

Aim Species distribution models (SDMs) are widely used to study geographic distributions of taxa in response to natural and anthropogenic environmental conditions. For a community, common approaches include fitting individual SDMs (iSDMs) to all taxa or directly modelling community properties such as richness. However, the parameters of iSDMs are difficult to identify for rare taxa, and community properties do not reveal taxon-specific responses. Individual models can be combined into a hierarchical multi-species distribution model (mSDM) that constrains taxon-specific parameters according to overarching community parameters, or a joint model (jSDM) in which interdependencies between taxa are jointly inferred. We compare how individual, hierarchical multi-species, and joint SDMs differ in quality of fit, explanatory power, and predictive performance, and analyze how these properties depend on the prevalence of taxa.

Taxa Presence-absence observations of 245 benthic macroinvertebrate taxa identified at a mixed taxonomic resolution.

Location 492 sites at rivers throughout Switzerland.

Methods Individual, hierarchical, and joint hierarchical generalized linear models (GLM) were developed for all taxa. Parameters were estimated using maximum likelihood estimation or Bayesian inference with Hamiltonian Markov chain Monte Carlo simulations. Predictive performance was assessed with cross-validation. In addition, the predicted family and species richness of the models was compared with a GLM for richness.

Results Individual models show a slightly higher quality of fit largely due to overfitting for rare taxa. The mSDM achieves a similar quality of fit and explanatory power, mitigates overfitting for rare taxa, and considerably improves predictive performance over the whole community. The joint models further improve the quality of fit, but decrease predictive performance and increase predictive uncertainty.

Main conclusions We show that even a relatively simple mSDM combines many of the analytical capabilities of iSDMs and improves predictive performance. Increasingly complex mSDMs and jSDMs provide additional analytical possibilities, but depending on the data and research questions, different levels of complexity may be appropriate.

Keywords: species distribution models, stacked individual models, hierarchical multi-species models, joint models, benthic macroinvertebrates, biodiversity, richness, predictive modelling

1 Introduction

Statistical species distribution models (SDMs) have become important tools for explaining and predicting the geographic distributions of taxa in response to key environmental conditions, yet ecologists are often confronted with complex community data (typically consisting of presence-absence or abundance observations at sampled sites) that is multivariate and sparse (i.e., containing a large number of taxa and many zeroes). Selecting an appropriate model given the data and research questions at hand can be daunting, given the variety of fundamentally different approaches in statistical species distribution modelling.

Perhaps the most common approach of applying a statistical SDM is to fit an individual SDM (iSDM) to a taxon, or to each taxon in a community to yield "stacked" iSDMs of the whole community (often with identical explanatory variables; Guisan and Rahbek, 2011). Stacked iSDMs quantify taxon-specific responses and can be used to estimate community properties such as species richness, although community properties can be modelled directly as a response variable in a richness model (RM; e.g., Kaelin and Altermatt, 2016). Both approaches have major shortcomings: unless taxon-specific iSDMs are used (with very few or no explanatory variables for rare taxa), studies applying stacked iSDMs often must exclude rare taxa due to the difficulty of identifying their parameters (Aitken et al., 2007; Elith et al., 2006; Guisan et al., 2006). Moreover, stacked iSDMs tend to overpredict species richness (e.g., Clark et al., 2014; Guisan and Rahbek, 2011), while RMs inherently ignore taxon-specific distributions. Another approach is to fit multiple iSDMs to a taxon and combine the models in an ensemble-based approach that can improve predictive performance and quantify predictive uncertainty due to different model structures. Although ensemble-based approaches have been shown to reduce predictive errors for rare taxa (e.g., Breiner et al., 2015; Lomba et al., 2010), the biological interpretation and comparison of multiple different models is more difficult relative to an individual model.

Rather than individually modelling the distributions of taxa independently of one another, more recent studies have applied multi-species and joint species distribution models (mSDMs and jSDMs, respectively) in which the responses of multiple taxa to environmental conditions are jointly inferred under one model (Latimer et al., 2009; Ovaskainen and Soininen, 2011; Warton et al., 2015b). Commonly implemented as an extended generalized linear model (GLM; see Guisan et al., 2002), an mSDM can quantify the responses of individual taxa according to a community response to the selected explanatory variables, leading to a hierarchical model structure. For instance, stacked iSDMs can be extended into a relatively simple hierarchical mSDM by introducing overarching community parameter distributions that effectively constrain the taxon-specific responses for each explanatory variable, similar to the model proposed by Ovaskainen and Soininen (2011). In a hierarchical mSDM, the response of the community and all individual taxa to the explanatory variables are quantified, allowing community properties such as richness to be derived as well. Moreover, hierarchical mSDMs can be extended with site- or sample-specific effects to quantify the effect of unknown explanatory variables not included in the model on the distributions of taxa, potentially guiding the selection of additional explanatory variables (Bolker et al., 2009; Warton et al., 2015a).

A joint model can include hierarchical structures similar to an mSDM, but additionally quantifies residual correlations between taxa after accounting for the explanatory variables included in the model (Pollock et al., 2014; Warton et al., 2015a). The residual correlations between taxa have been used to explore and test hypotheses of patterns between distributions of taxa arising due to biotic interactions or unknown explanatory variables, and to quantify associations between taxa under varying environmental conditions (e.g., Tikhonov et al., 2017) or spatial scales (e.g., Ovaskainen et al., 2016a). Studies applying jSDMs have emphasized that the residual correlations between taxa do not directly quantify biotic interactions, and should be interpreted with caution (Clark et al., 2014; Dormann et al., 2018). For example, a positive correlation between two species may be due to mutualism or competition

(e.g., due to high niche overlap; see Clark et al., 2014), or due to an explanatory variable not included in the model (Ovaskainen et al., 2010; Tikhonov et al., 2017; Warton et al., 2015a). Joint models have been applied to a variety of communities (e.g., Pollock et al., 2014), including waterbirds and bryophytes (e.g., Ovaskainen et al., 2017), trees (e.g., Clark et al., 2014), butterflies (e.g., Ovaskainen et al., 2016b), and marine fish (e.g., Thorson et al., 2016). Comparative studies have found that relative to individual modelling approaches, equivalent hierarchical multi-species and joint model structures improve predictive performance for the overall community (e.g., Clark et al., 2014; Hui et al., 2013; Ovaskainen and Soininen, 2011; Taylor-Rodríguez et al., 2017), and quantify community properties such as species richness (e.g., Clark et al., 2014; Taylor-Rodríguez et al., 2017).

Despite many applications, few studies have conducted a comprehensive methodological comparison of the performance of stacked individual, hierarchical multi-species, and joint species distribution models for community modelling and prediction, or for modelling community properties such as richness (e.g., Clark et al., 2014; Hui et al., 2013; Ovaskainen et al., 2016b; Taylor-Rodríguez et al., 2017; Tikhonov et al., 2017). In this paper, we compare the performance of equivalent individual and joint modelling approaches, incrementally extending an individual model into a hierarchical multi-species and finally into a joint model that quantifies both the effects of environmental conditions on each taxon and the residual correlations between taxa. Based on these models, we explore and analyze differences in model performance using an extensive dataset of benthic stream macroinvertebrate communities in Switzerland with a wide range in prevalence (i.e., the proportion of samples in which a taxon was present) and many rare taxa, specifically emphasizing how model performance is affected by the prevalence of the taxa. While comparing these models we pose the following questions:

1. To what extent and why do individual models, a hierarchical multi-species model (of varying complexity), and a joint model differ in quality of fit, explanatory power, and predictive performance for the overall community? How does model performance depend on the prevalence of a taxon in a community with many rare taxa?

2. To what extent can we quantify the residual correlations between taxa in a joint model, and to what extent does this improve model performance with respect to a hierarchical multi-species model?

3. How well do hierarchical multi-species and joint models predict richness when compared to modelling richness directly?

We address our research questions by performing a systematic variable selection procedure using iSDMs (i.e., stacked individual GLMs) to select a set of explanatory variables that optimize the mean predictive performance for taxa with identifiable parameters. We compare the quality of fit during calibration and predictive performance under cross-validation of individual models to an equivalent hierarchical multi-species model (i.e., an mSDM). We then extend the mSDM to include correlated hierarchical parameters, site-specific effects, and residual correlations between taxa (parameterized with latent variables) to obtain a hierarchical joint generalized linear model (i.e., a jSDM). We aim to analyze how these model

elements affect model performance, the quantification of community- and taxon-specific responses, and richness.

We hypothesize that (1) the iSDMs will exhibit a slightly better quality of fit during calibration than an mSDM with an equivalent model structure, but will be prone to overfitting for increasingly rare taxa due to unconstrained parameter estimates. The hierarchical structure of the mSDM should achieve a similar quality of fit, constrain the parameters for individual taxa, mitigate overfitting, and improve predictive performance for the community. We expect that (2) relative to an mSDM, a joint model quantifying residual correlations between taxa should further improve predictive performance due to the role of biotic interactions or joint responses to unknown explanatory variables in driving community composition. Following from (2) we expect that (3) a joint model should improve performance for predicted richness relative to an RM or mSDM.

2 Methods

2.1 Study Area and Data

We used data from the program 'Biodiversity Monitoring in Switzerland' (BDM; BDM Coordination Office, 2009). It consists of semi-quantitative abundance observations of macroinvertebrates based on Swiss multi-habitat sampling protocols for stream assessments (Stucki, 2010). A total of 580 samples were taken at 492 sites located throughout Switzerland using the Swiss topographic grid and sampled between 2011 and 2015 during March and July, depending on the elevation of the sites (Figure S1.1). For the purposes of this study, we used presence-absence observations of 245 benthic macroinvertebrate taxa at a mixed taxonomic resolution (based on the highest level of identification, with species resolution only for EPT[1] orders).

2.2 Model Definitions

In defining the individual, hierarchical multi-species, and joint models below in a stepwise approach, we identified six distinct models in our analysis to quantify how various aspects of model performance depend on increasing model complexity. The following indices are used throughout the paper:

Sites:	$i \in \{1, \dots, I\}$
Samplings at site i:	$t_i \in \{1, \dots, T_i\}$
Taxa ("species"):	$j \in \{1, \dots, J\}$
Explanatory variables:	$k \in \{1, \dots, K\}$
Latent variables	$l \in \{1, \dots, L\}$

For a presence-absence response where observations are $y_{it_ij} = 1$ if a taxon is observed as present and $y_{it_ij} = 0$ if the taxon is not observed as present, we apply a generalized linear

---

[1]Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies).

model with a logistic link function to each taxon to obtain probabilistic predictions that the random variable $Y_{it_ij} = 1$ (see Guisan et al., 2002). The probability of occurrence (including the probability of detection) $P(Y_{it_ij} = 1 \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}})$ given the explanatory variables $\mathbf{x}$, and the taxon-specific parameters $\boldsymbol{\alpha}^{\text{taxa}}$ and $\boldsymbol{\beta}^{\text{taxa}}$ is equal to the link function (Equation 2) evaluated at a linear combination of the explanatory variables and parameters (Equation 1). The explanatory variables shift the linear predictor z (Equation 1) that, through the logistic link function, yields a probability of occurrence (Equation 2).

$$z_{it_ij} = \alpha_j^{\text{taxa}} + \sum_{k=1}^{K} x_{it_ik}\beta_{kj}^{\text{taxa}} \quad , \tag{1}$$

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

The observations are assumed to be Bernoulli distributed with $P\left(y_{it_ij} \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}\right)$ equal to $P\left(Y_{it_ij} = 1 \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}\right)$ if a taxon is present at a site ($y_{it_ij} = 1$) or $1 - P\left(Y_{it_ij} = 1 \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}\right)$ if a taxon is absent ($y_{it_ij} = 0$). When fitting an iSDM to each taxon (i.e., stacked iSDMs) in the community, the joint probability of the observations of all taxa given the model structure, explanatory variables, and parameters is

$$P\left(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}\right) = \prod_{i=1}^{I}\prod_{t_i=1}^{T_i}\prod_{j=1}^{J} P\left(y_{it_ij} \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}\right) \quad . \tag{3}$$

In applying an identical iSDM to each taxon, the parameters for rare taxa may be difficult to identify. This limitation is addressed in the mSDM by assuming that the taxon-specific parameters α_j^{taxa} and β_{kj}^{taxa} are distributed according to and constrained by the corresponding community distributions α^{comm} and β_k^{comm} (Ovaskainen and Soininen, 2011)[2]. 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 probability densities formulated as a probability network (Figure 1).

---

[2]If the parameters vary strongly across taxa, the community distributions would be very wide and thus would not provide much information for rare taxa.
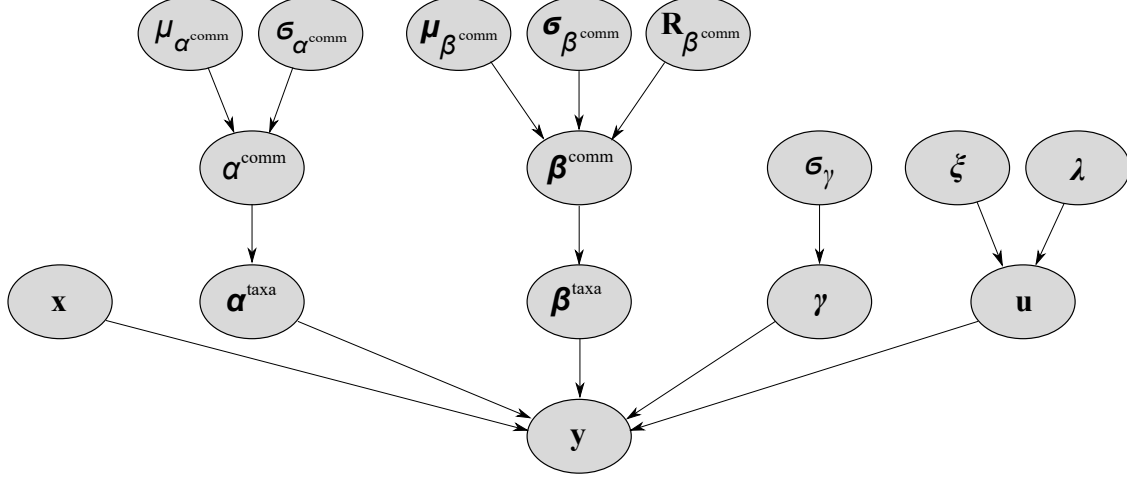
Figure 1: Network representation of conditional probability distributions in the hierarchical multi-species and joint models. The nodes $\mu_{\alpha^{\text{comm}}}, \sigma_{\alpha^{\text{comm}}}$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}^{\text{comm}}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}^{\text{comm}}}$ are the means and standard deviations parameterizing their respective community distributions α^{comm} and $\boldsymbol{\beta}^{\text{comm}}$. The nodes $\boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}$ are taxon-specific parameters constrained by and jointly inferred with their respective community distributions $\alpha^{\text{comm}}, \boldsymbol{\beta}^{\text{comm}}$ from the model inputs (i.e., explanatory variables) $\mathbf{x}$ and model output $\mathbf{y}$. The conditional probability densities extending the mSDM are discussed in-text below, including the correlation matrix of the community parameters ($\mathbf{R}_{\boldsymbol{\beta}^{\text{comm}}}$), the site effect ($\boldsymbol{\gamma}$ and its parameter σ_{γ}), and the random effect $\mathbf{u}$ (given by the latent variables $\boldsymbol{\xi}$ and factor loadings $\boldsymbol{\lambda}$).

Assuming normal distributions for the community distributions α^{comm} and β_k^{comm} with corresponding parameters $\mu_{\alpha^{\text{comm}}}, \sigma_{\alpha^{\text{comm}}}$ and $\mu_{\beta_k^{\text{comm}}}, \sigma_{\beta_k^{\text{comm}}}$, the joint prior probability distribution of the parameters and outcomes in the mSDM is

$$
\begin{aligned}
P\left(\mathbf{y}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}, \mu_{\alpha^{\text{comm}}}, \sigma_{\alpha^{\text{comm}}}, \boldsymbol{\mu}_{\boldsymbol{\beta}^{\text{comm}}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}^{\text{comm}}} \mid \mathbf{x}\right) &= \\
P(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}) \times \prod_{j=1}^{J} \left(f_{\text{N}}(\alpha_j \mid \mu_{\alpha^{\text{comm}}}, \sigma_{\alpha^{\text{comm}}}) \times f_{\beta}(\beta_j \mid \boldsymbol{\mu}_{\boldsymbol{\beta}^{\text{comm}}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}^{\text{comm}}}, \mathbf{R}_{\boldsymbol{\beta}^{\text{comm}}}) \right) & \quad (4) \\
\times f\left(\mu_{\alpha^{\text{comm}}}, \sigma_{\alpha^{\text{comm}}}, \boldsymbol{\mu}_{\boldsymbol{\beta}^{\text{comm}}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}^{\text{comm}}}\right) & \quad ,
\end{aligned}
$$

where f_{N} is the density of the one-dimensional normal distribution and f is the density of the joint prior of the means and standard deviations of the community parameters (f is assumed to be the product of normal distributions for the means and lognormal distributions for the standard deviations) and f_{β} is the density of a multivariate normal distribution:

$$
f_{\beta}(\beta_j \mid \boldsymbol{\mu}_{\boldsymbol{\beta}^{\text{comm}}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}^{\text{comm}}}, \mathbf{R}_{\boldsymbol{\beta}^{\text{comm}}}) = f_{\text{N}_K}(\beta_j \mid \boldsymbol{\mu}_{\boldsymbol{\beta}^{\text{comm}}}, \text{diag}(\boldsymbol{\sigma}_{\boldsymbol{\beta}^{\text{comm}}}) \cdot \mathbf{R}_{\boldsymbol{\beta}^{\text{comm}}} \cdot \text{diag}(\boldsymbol{\sigma}_{\boldsymbol{\beta}^{\text{comm}}})) \quad . \quad (5)
$$

In the simplest mSDM, the correlation matrix $\mathbf{R}_{\beta\text{comm}}$ is assumed to be diagonal to reflect the assumption that the community parameters are independent of each other. Considering correlated community parameters may improve the identifiability of the taxon-specific parameters and improve the quality of fit and predictive performance, and reveal dependencies among community responses to the explanatory variables. To do so, we can assume an LKJ prior (Lewandowski et al., 2009) for the correlation matrix given by

$$f(\mathbf{R}_{\beta\text{comm}}) \propto \det(\mathbf{R}_{\beta\text{comm}})^{\eta-1} \quad , \tag{6}$$

as an additional factor to Equation 4 (with $\eta = 2$).

The addition of a site- or sample-specific effect can facilitate the selection of additional explanatory variables or be used to identify non-linear responses to environmental conditions (by comparing the site effect with potential explanatory variables). In the multi-species and joint models defined here, we include a site effect γ_i that increases or decreases the predicted probabilities for all taxa at each site due to site-specific explanatory variables not included in the model:

$$z_{it_ij} = \alpha_j^{\text{taxa}} + \sum_{k=1}^{K} x_{it_ik}\beta_{kj}^{\text{taxa}} + \gamma_i \quad . \tag{7}$$

The site effect is assumed to be normally distributed with mean zero and standard deviation σ_γ, with a lognormal prior for σ_γ. This results in an additional factor for Equation 4:

$$f(\gamma_i, \sigma_\gamma) = f_{\text{N}}(\gamma_i \mid 0, \sigma_\gamma) \times f_{\text{LN}}(\sigma_\gamma) \quad . \tag{8}$$

To extend the mSDM (with correlated community parameters and a site effect) into a joint model, residual correlations between taxa can be modelled as a multivariate random effect over the observational units (e.g., sites or samples) (Warton et al., 2015a). When the number of taxa is small relative to the number of samples, the parameters of the multivariate random effect may be inferred from the data based on an unstructured covariance matrix (Warton et al., 2015a). However, the number of parameters quickly increases for communities with a large number of taxa. This limitation can be mitigated using dimensionality reduction techniques such as Dirichlet processes (e.g., Taylor-Rodríguez et al., 2017) or latent variables (e.g., Letten et al., 2015; Ovaskainen et al., 2016a; Tikhonov et al., 2017; Warton et al., 2015a). Latent variables are particularly appealing as residual correlations between taxa are induced by taxon-specific parameters or "factor loadings" that respond to a site- or sample-specific latent variable (i.e., the latent variable is a random effect). Multiple latent variables can be included in a joint model to further improve the identifiability of the true correlation structure among taxa, but at the cost of increasing the number of parameters (Warton et al., 2015a). In our model, we introduce a random effect u_{it_ij} to the linear predictor

$$z_{it_ij} = \alpha_j^{\text{taxa}} + \sum_{k=1}^{K} x_{it_ik}\beta_{kj}^{\text{taxa}} + \gamma_i + u_{it_ij} \quad . \tag{9}$$

8

The dimensionality of the random effect $u_{it_i j}$ can be reduced with a latent variable parameterization defined as

$$u_{it_i j} = \sum_{l=1}^{L} \xi_{it_i l} \lambda_{lj} \quad , \tag{10}$$

with

$$f(\xi_{it_i l}) = f_N(\xi_{it_i l} \mid 0, 1) \tag{11}$$

and

$$f(\lambda_{lj} \mid \sigma_{\lambda_l}) = f_N(\lambda_{lj} \mid 0, \sigma_{\lambda_l}) \tag{12}$$

as additional factors to Equation 4. The latent variables $\xi_{it_i l}$ can be interpreted either as unknown explanatory variables or as the main axes of covariation in occurrences across taxa (Warton et al., 2015a). The taxon-specific factor loadings λ_{lj} induce correlations between the predicted probabilities of occurrence among taxa in each sample in response to the latent variables (Warton et al., 2015a). The covariance matrix quantifying associations between taxa is given by the residual covariance matrix

$$\boldsymbol{\Sigma} = \boldsymbol{\lambda}^T \boldsymbol{\lambda} \quad . \tag{13}$$

To test how model performance depends on various model extensions (e.g., including correlated community parameters, site effects, and latent variables) that increase model complexity, we selected six distinct individual, hierarchical multi-species, and joint models for testing (Table 1). Throughout the text, we will refer to these models using a consistent set of abbreviations defined in Table 1.

Table 1: Abbreviations and extensions of the individual, hierarchical multi-species, and joint models, with models ordered by increasing number of parameters and model extensions highlighted (in grey).

In-text abbreviation	Model structure	Number of parameters	Extensions		
			Correlated (C) / uncorrelated (U) $\boldsymbol{\beta}^{\mathrm{comm}}$	With (T) / without (F) site effect $(\boldsymbol{\gamma})$	Number of latent variables $(\boldsymbol{\xi}, \boldsymbol{\lambda})$
iSDM	iSDM	1715	U	F	0
UF0	mSDM	1729	U	F	0
CF0	mSDM	1765	C	F	0
CT0	mSDM	2258	C	T	0
CT1	jSDM	3083	C	T	1
CT2	jSDM	3908	C	T	2

Note: the number of parameters are calculated based on model calibration to the entire dataset with six explanatory variables.

2.3   Model Performance

Each model was first calibrated using the entire dataset to comprehensively demonstrate differences in quality of fit and how these differences depend on the model structures and prevalence of taxa in the community. We included rare taxa (i.e., with too few presence points; see Figure S1.1) when calibrating all models, including the iSDMs to analyze how hierarchical and joint model structures affect model performance. The quality of fit, explanatory power of the input variables, and predictive performance of each taxon in the individual, multi-species, and joint model is based on the deviance (a statistic equivalent to the residual sum of squares under the assumption of normality) over all observations and given as

$$d_j = -2 \sum_{i=1}^{I} \sum_{t_i=1}^{T_i} \log \left(P \left(y_{it_ij} \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}} \right) \right) \quad , \tag{14}$$

where d_j is the taxon-specific deviance over all samples based on the probabilities of occurrence predicted by the null or proposed model (Equation 14). Under the null model, the taxon-specific intercept α_j^{taxa} corresponds to the prevalence of a taxon and its parameters β_{kj}^{taxa} are set to zero. The deviance can be divided by the number of samples for a taxon to obtain a standardized deviance that is comparable among taxa with different sample sizes. We chose to use the deviance to quantify model performance rather than the classification statistics commonly used in studies applying SDMs (e.g., Mouton et al., 2010) to avoid imposing different classification thresholds on the probabilities of occurrence of the taxa, and to ensure comparable model predictions among taxa that vary widely in their prevalence.

For the iSDM and mSDM UF0, the explanatory power or D_j^2 (a statistic similar in interpretation to the R^2 of a linear model) of the selected explanatory variables for each taxon is measured as the proportion of deviance in the null model that is reduced by the proposed model (i.e., including the explanatory variables; Guisan and Zimmermann, 2000):

$$D_j^2 = \frac{d_j^{\text{null}} - d_j^{\text{proposed}}}{d_j^{\text{null}}} \quad . \tag{15}$$

To assess the geographic distribution of the quality of fit for each taxon in the iSDMs and mSDM UF0, the maximum likelihood and maximum posterior parameter values were propagated through the respective models to obtain point estimates of the probabilities of occurrence for all taxa at each sample. The uncertainty in the probabilities of occurrence of each taxon in UF0 was quantified by propogating the posterior sample through the model and calculating the 5th and 95th quantiles of the predicted probabilities.

The predictive performance of the individual, hierarchical multi-species, and joint models was assessed using k-fold cross-validation. The samples were randomly divided into $k = 3$ subsamples of nearly equal size. Each model was calibrated using each combination of

$k-1$ subsamples and used to predict the probability of occurrence for each taxon in the independent subsample. We chose $k=3$ due to the computationally intensive inference of the multi-species and joint models, and to avoid deteriorating the predicted probabilities of occurrence in the independent subsample. The predictive performance for each taxon was quantified as the mean standardized deviance of k subsamples.

The predictive performance of the hierarchical multi-species and joint models with site effects and latent variables was quantified for all taxa using both the maximum posterior parameter values and a sample of the joint posterior parameter distribution. For the mean standardized deviance at the maximum posterior, the site effect and latent variables were set to zero to enable a comparison with the individual models and to analyze how the site effect and latent variables (i.e., in CT0, CT1, and CT2) shift the taxon-specific parameters with respect to simpler hierarchical multi-species models (i.e., UF0, CF0).

The predictive uncertainty of the hierarchical multi-species and joint models was compared by propagating the joint posterior parameter distribution through the models to obtain predicted probabilities of occurrence and the mean standardized deviances of taxa during prediction for independent data. During predictions for independent data, the site effect was simulated by randomly sampling from its normal distribution (Equation 8) while the latent variables were simulated by randomly sampling from a multivariate normal distribution with a zero mean and a standard deviation given by the residual covariance matrix (Equation 13).

2.4 Parameter Inference & Model Implementation

The taxon-specific parameters $\alpha_j^{\text{taxa}}, \beta_{kj}^{\text{taxa}}$ in the individual models were identified by maximum likelihood estimation. Numerically, parameter estimates were obtained with an iterative weighted least squares (IWLS) algorithm in the *glm* function of the R statistical computing environment (R Core Team, 2018). The null model parameters were used as starting values for the IWLS algorithm. If the IWLS algorithm produced parameters where the proposed model deviance was greater than the null model deviance (see equation Equation 14), a more robust optimization method by Nelder and Mead (1965) (implemented in the *optim* function in R) was applied to identify the maximum likelihood solution. The use of efficient maximum likelihood estimation allowed us to obtain parameter estimates for a very large number of combinations of potentially important explanatory variables for variable selection (see subsection 2.6).

The parameters of the hierarchical multi-species and joint models were estimated by Bayesian inference. Wide priors were used for the means and standard deviations of the community distributions to compare the maximum posterior parameter estimates to the maximum likelihood parameter estimates for the individual models, and to clearly distinguish the effect of the simplest hierarchical multi-species model (i.e., UF0) structure on the taxon-specific parameters. The joint posterior probability distributions of the model parameters were sampled by doing Bayesian inference with a Hamiltonian Markov chain Monte Carlo (HMC; see Duane et al., 1987, Brooks et al., 2011) algorithm implemented in Stan and accessed through the R package *rstan* (Stan Development Team, 2016; see Appendix S1 subsection S1.10 for

scripts containing Stan model definitions that correspond to the mSDMs and jSDMs summarized in Table 1). To improve inference of the correlation coefficients of the correlated community parameters, we used a Cholesky decomposition of the correlation matrix $\mathbf{R}_{\beta^{\mathrm{comm}}}$ as provided in Stan. Convergence issues for joint models with multiple latent variables (e.g., CT2) are detailed in subsection S1.8 of Appendix S1.

2.5 Community Properties & Taxon-Specific Responses

Due to the mixed taxonomic resolution of the dataset, we calculated two metrics of richness to compare model predictions to the observations: species richness for EPT taxa identified at the species level, and family richness with all taxa aggregated to the family level or lower. The predicted family and species richness was obtained during k-fold cross-validation of a model for richness as a directly modelled output (i.e., a richness model, RM), and calculated from the taxonomically resolved output of the hierarchical mSDM and jSDM (see subsection 2.3). For the mSDM and jSDM, the predicted richness at each sample was obtained by sampling from the model including the Bernoulli distribution for presence-absence. In the RM, we assume a negative binomial GLM with a log-mean richness given by the linear predictor (Equation 16, Equation 17):

$$P(r_{it_i} \mid \mathbf{x}, \alpha, \boldsymbol{\beta}, \theta) = P_{\mathrm{NB}}(r_{it_i} \mid \mu_{it_i}(\mathbf{x}, \alpha, \boldsymbol{\beta}), \theta) \quad , \tag{16}$$

$$\log(\mu_{it_i}) = \alpha + \sum_{k=1}^{K} x_{it_ik}\beta_k \quad , \tag{17}$$

where r_{it_i} is the richness at the sampled site i at time t_i (i.e., the number of taxa) and θ is an additional parameter to control the variance of the negative binomial distribution ($\sigma^2 = \mu + \mu^2/\theta$) (Equation 16). The dispersion parameter θ is estimated from the data in addition to the parameters α and $\boldsymbol{\beta}$ (Equation 16). This model is similar to the model used by Kaelin and Altermatt (2016), though we use different explanatory variables and use a negative binomial rather than a Poisson distribution to describe the variance more realistically. The richness model was fitted (one parameter α, K parameters β_k, and one parameter θ) using the function *glm.nb* in the R package *MASS* (Venables and Ripley, 2002).

The predicted and observed family and species richness for the RM and mSDM UF0 was compared based on their respective maximum likelihood and maximum posterior parameter estimates. The predictive performance and uncertainty of the hierarchical multi-species (UF0-CT0) and joint models (CT1, CT2) was compared by propagating a subsample of the posterior distribution through each model during k-fold cross-validation (see subsection 2.3).

Taxon-specific responses in the individual models and the mSDM UF0 were obtained by plotting the probability of occurrence (for the entire dataset) in response to the explanatory variables. The posterior parameter distributions underlying the model predictions and the hierarchical structure of the mSDM UF0 were both visualized by identifying statistically significant positive and negative responses at a 5% level based on the respective 5th and 95th quantiles of the marginal posterior taxon-specific parameter distributions ($\beta_{kj}^{\mathrm{taxa}}$) for all taxa.

2.6 Selection of Explanatory Variables

Using a broad range of potential variables chosen based on *a priori* knowledge, we applied stacked iSDMsin an exhaustive search procedure (i.e., similar to a best subsets or "brute force" approach; see James et al., 2013) to identify a set of explanatory variables that maximizes the predictive performance for taxa with a minimum number of occurrences (due to the difficulty of identifying the parameters for rare taxa). Potential explanatory variables were identified on the basis of expert knowledge and existing literature, with an emphasis on direct, mechanistic explanatory variables reflecting natural and anthropogenic impacts on stream invertebrate communities in Switzerland (Austin, 2007; Guisan and Zimmermann, 2000) (Table 2; see Table S1.1 and Table S1.2 for detailed descriptions of each variable).

Table 2: Potential explanatory variables of stream invertebrate communities in Switzerland (selected variables in bold)

Description	Abbreviation (units)
Mean maximum morning summer stream temperature predicted from a linear model based on catchment area and mean catchment elevation	**Temp** (°C)
Mean annual stream flow velocity estimated from spatial data	**FV** (m/s)
Riparian agriculture (A) or forest (F) cover within a buffer distance from stream	A10m, A100m, A1km, F10m, **F100m**, F1km (%)
Livestock unit density: cattle equivalent (CE) units of livestock per square kilometre of catchment area	**LUD** (CE/km^2)
Insecticide application rate: sum of the products of the fraction of croplands in the catchment and respective mean annual spray treatments	**IAR** (unitless)
Proportion of urban and transport-related land use within the catchment	Urban (%)
Forest-river intersection: proportion of upstream river length in the catchment intersected by forest cover	FRI (%)
Proportion of forest cover within the catchment	Forest (%)
Evaluation of width variability of stream channel (0 = bad, 1 = very good)	WV (unitless)
Evaluation of stream bed modification (0 = bad, 1 = very good)	BM (unitless)
Evaluation of overall stream morphology (0 = bad, 1 = very good)	Morph (unitless)
Agricultural (A) or forest (F) land cover in 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 (%)
Wastewater discharge as a proportion of stream discharge	WW (%)
Residual flow: fraction of residual stream discharge from upstream hydroelectric dams	RF (unitless)

Note: 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 while still maintaining parameter estimates relative to the units of the explanatory variables. Temp and FV include a quadratic transformation to identify taxa with mid-range preferences for temperature and flow velocity (i.e., **Temp^2**, FV^2). For a complete description of each variable, see Table S1.1 and Table S1.2.

The exhaustive search procedure uses k-fold cross-validation to quantify the mean standard-

ized deviance for each taxon based on independent predictions (see subsection 2.3) for a full subset of models of varying complexity (i.e., between five and eleven parameters). For any number of parameters p, models containing all possible combinations of p potential explanatory variables were constructed. The number of proposed models was constrained by only considering models that included stream temperature (Temp), flow velocity (FV), and the insecticide application rate (IAR) as explanatory variables, and excluding models with collinear explanatory variables (i.e., a Pearson correlation greater than 0.6 or less than -0.6). The predictive performance of each potential model for the whole community was calculated as the mean standardized deviance during k-fold cross-validation of all taxa with identifiable parameters (see subsection 2.3; e.g., taxa in BDM occurring in less than 10% of the samples were excluded due to the poor identifiability of their parameters (Figure S1.8) and consistently poor predictive performance).

3 Results

3.1 Selection of Explanatory Variables

The exhaustive search procedure shows that simpler individual models (i.e., with fewer parameters) have a higher deviance during calibration but maintain a relatively similar predictive error, demonstrating that simpler models produce a more stable predictive performance while more complex models tend to overfit (Figure 2). However, multiple models consisting of between approximately six and nine parameters achieve a higher predictive performance over the community relative to models with fewer or a greater number of parameters.
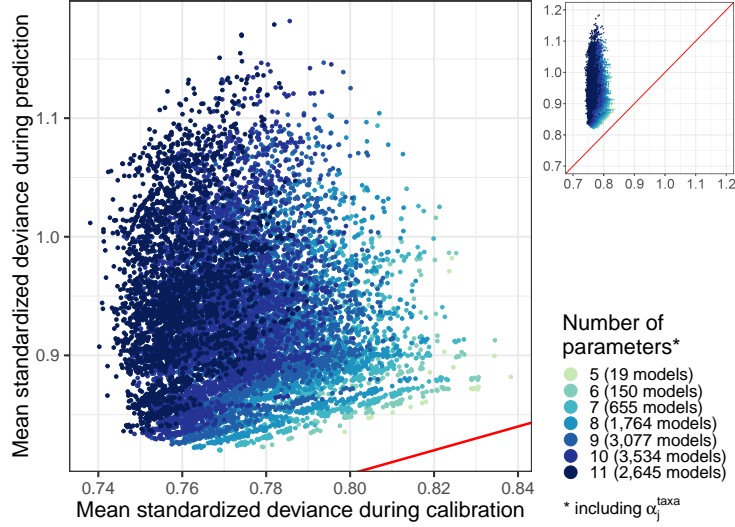
Figure 2: Mean standardized deviance during calibration and prediction of potential individual models with between five and eleven parameters, with a 1:1 reference line (red). Each point represents the predictive performance of a potential model for taxa with identifiable parameters (see subsection 2.3 and Figure S1.8). The inset figure (top right) shows the same data with rescaled axes.

The top ten models with the highest mean predictive performance among the datasets show a generally consistent set of explanatory variables, including multiple agricultural (IAR, LUD) and forest-related explanatory variables (F10m, F100m, or FRI; see Table S1.3). The explanatory variables of the iSDMs with the lowest mean standardized deviance during prediction were selected as the basis for further analysis (see Table 2 with selected variables shown in bold; see Figure S1.2 for the geographic distribution of the selected explanatory variables).

3.2   Quality of Fit for the Full Dataset

Calibrating the iSDMs and the simplest mSDM (UF0) to the entire dataset with identical explanatory variables reveals that the dependence of the quality of fit (i.e., the standardized deviance) on the prevalence of the taxa differs based on the model structure (Figure 3a). The null model (i.e., including an intercept α_j^{taxa} and with $\beta_{kj}^{\text{taxa}} = 0$) deviance is lowest for rare and very common taxa (given that a good fit does not require a high explanatory power), and highest for taxa of intermediate prevalence where explanatory variables are needed to distinguish modelled predictions in samples where the taxa are present or absent. The lower standardized deviance values of the iSDMs suggest a slightly higher quality of fit for increasingly rare taxa than the UF0 model, however these differences become negligible for intermediately prevalent taxa with distributions that are well explained by the selected variables (Figure 3a).
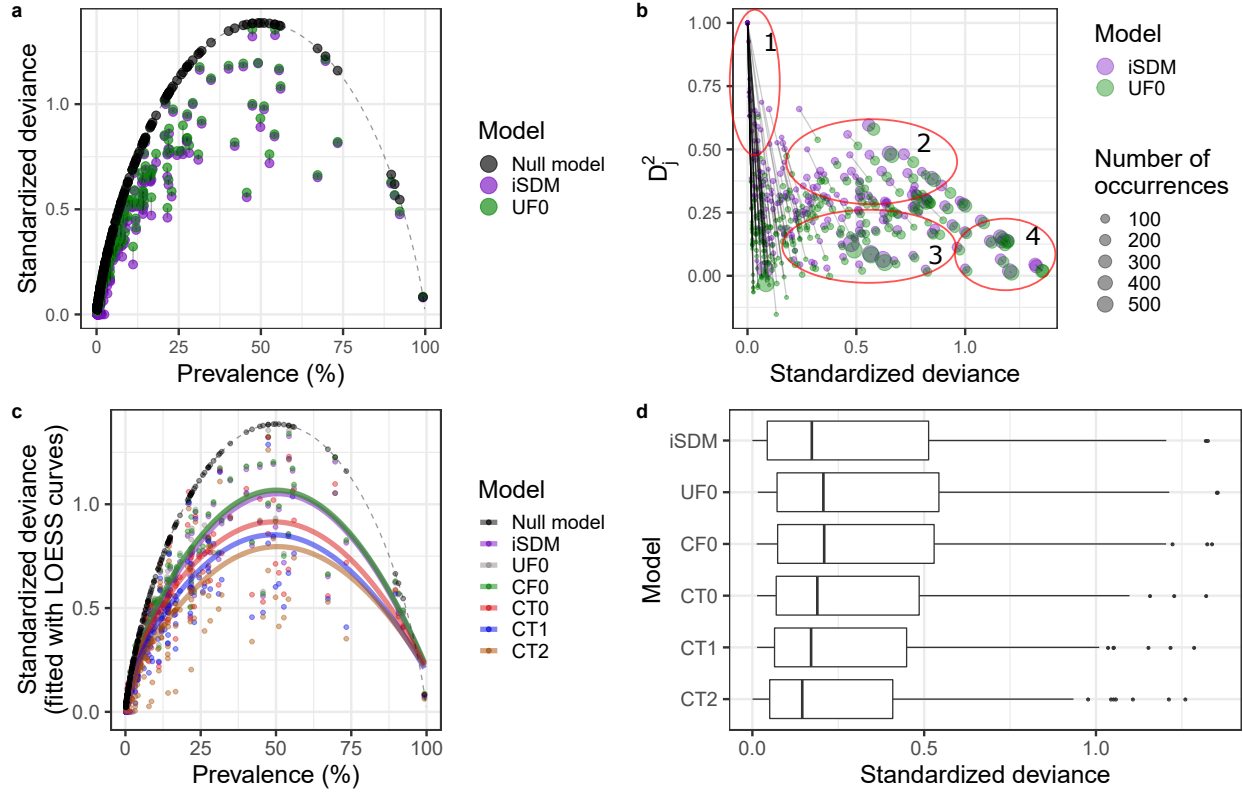
Figure 3: **Top**: Quality of fit (standardized deviance) during calibration to the entire dataset and explanatory power (D_j^2) of the iSDMs and the mSDM UF0, including (a) the standardized deviance of the null model (i.e., intercept only; black points) and proposed models versus the prevalence of each taxon (i.e., the proportion of samples where the taxon was observed to be present), and (b) the explanatory power of the variables versus the standardized deviance for each taxon, with the length of the lines visualizing differences between the models (approximate regions of interest are circled in red and described in-text). **Bottom**: comparisons of the quality of fit of the iSDMs and extended model structures, including (c) the standardized deviance values for each proposed model (and the null model given by black points and dashed line; note that the LOESS curves for UF0 and CF0 overlap) versus the prevalence of each taxon, and (d) the distribution of standardized deviance for taxa in the proposed models.

Differences in the explanatory power (D_j^2) of the iSDMs and UF0 are in turn dependent on the quality of fit of the proposed model (Figure 3b). For rare taxa in region (1), the individual models produce a nearly perfect quality of fit and explanatory power ($d_j \approx 0$, $D_j^2 \approx 1$) of the selected variables, while UF0 produces a relatively lower explanatory power. Taxa in region (2) have an intermediate prevalence and a reasonable explanatory power of the selected variables ($0.20 < D_j^2 < 0.60$) with an intermediate quality of fit. Taxa in region (3) display a similar quality of fit and yet are poorly explained by the selected variables. Region (4)

includes common taxa with a poor quality of fit and low explanatory power of the selected variables in both models (Figure 3b).

Compared to the simplest mSDM UF0, introducing correlated community parameters (i.e., CF0) has no significant effect on the quality of fit (Figure 3cd). However, incorporating site effects with CT0 produces a significant improvement in the quality of fit over the whole community while the joint models (i.e., including one or two latent variables: CT1, CT2) yield similar improvements (Figure 3cd).

3.3 Predictive Performance during Cross-Validation

During cross-validation, the iSDMs of multiple taxa have a low mean standardized deviance during calibration that increases significantly during prediction (i.e., for samples not used to calibrate the model), indicating a tendency to overfit (especially among rare taxa; Figure 4ab). The standardized deviance for each taxon in the mSDM UF0 remains similar during calibration and prediction, demonstrating that the model consistently generalizes to independent data and mitigates overfitting for rare taxa. The addition of correlated community parameters with CF0 yields no significant improvement in predictive performance (Figure 4b), while the site effects (CT0) and latent variables (CT1, CT2) lead to a decrease in the mean standardized deviance during calibration (i.e., improving the quality of fit) and increased and wider distributions of predictive error (Figure 4c). Joint models including three or four latent variables were also tested (with the results not shown here), but led only to further decreases in predictive performance.
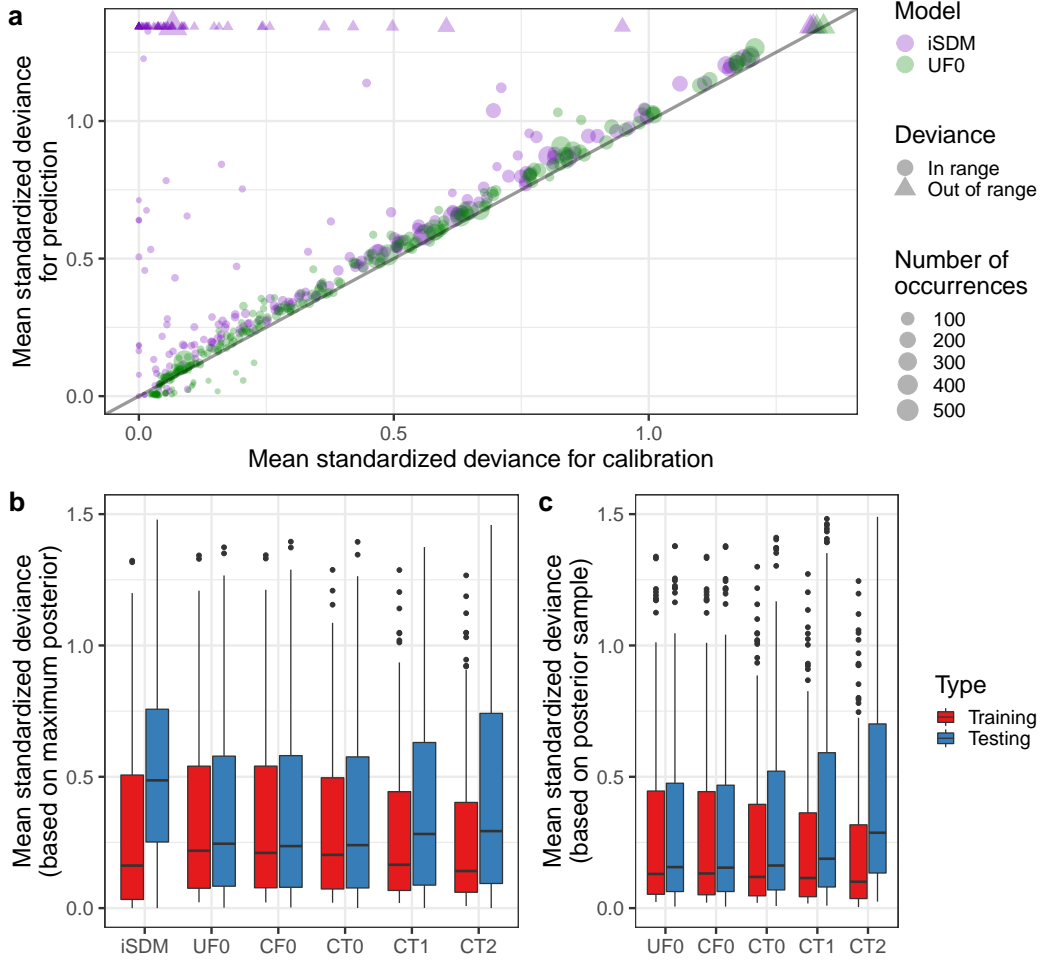
Figure 4: Predictive performance during cross-validation for taxa in the individual, hierarchical multi-species, and joint models, including (a) the mean standardized deviance during calibration and prediction for taxa in the individual models versus the simplest hierarchical multi-species model (UF0) based on maximum likelihood and maximum posterior parameter estimates respectively(including a 1:1 reference line), (b) the mean standardized deviance (over the cross-validation folds) at the maximum posterior for taxa in each model, and (c) the mean standardized deviance values based on the posterior sample (i.e., a sample of the joint posterior distribution) of the hierarchical multi-species and joint models (UF0-CT2). Note that (a) shows predictive performance during calibration and prediction while (b) and (c) shows the predictive performance only during prediction (i.e., for independent testing data). In (b), $\gamma = 0$ and $u = 0$ during predictions for models CT0, CT1, and CT2.

3.4   Predicted Richness during Cross-Validation

The mSDM UF0 and the richness model produce similar predictions for family and species richness during cross-validation, with a bias at very low and high observed species richness

due to over- and under-predictions, respectively (Figure 5, Figure S1.9). Propagating the joint posterior distribution through the mSDMs and joint models shows that species richness remains more biased than family richness, although the mean posterior family richness has more variability across all models. Over the posterior sample, the simpler models (UF0, CF0) generally have a lower uncertainty in predicting richness while the mSDM with site effects (CT0) and joint models (CT1, CT2) have relatively larger predictive uncertainties. However, the predictive performance of the mean posterior richness appears to become more narrow and biased when extending the simplest mSDM (UF0) to the joint model with two latent variables (CT2).
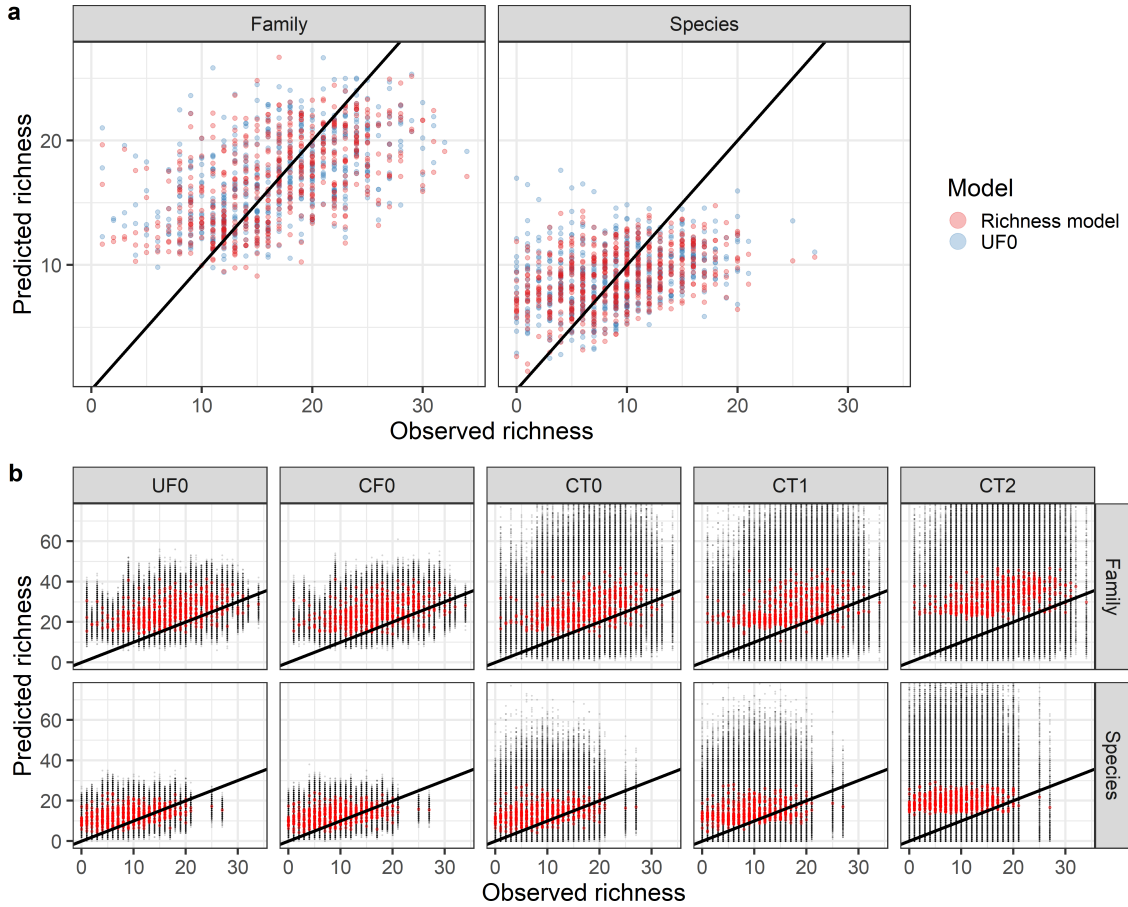
Figure 5: Predicted versus observed family and species richness at each sample (of a site) during cross-validation of the richness model, hierarchical multi-species and joint models, including (a) the mean predictions of the richness model and the hierarchical multi-species model (UF0) based on their respective maximum likelihood estimates and maximum posterior parameter estimates, and (b) the hierarchical multi-species and joint models, with the posterior predicted richness distribution (black points) and the mean posterior predicted richness (red points) at each sample.. One-to-one reference lines are included.

3.5   Community & Taxon-Specific Responses

The hierarchical structure of the simplest mSDM (UF0) allows rare taxa with a high uncertainty in their taxon-specific responses (i.e., the marginal posterior taxon-specific parameters β_{kj}^{taxa}) to "borrow" information from the community response (i.e., the posterior community parameters β_k^{comm}), resulting in improved predictive performance relative to the iSDMs (e.g., Figure 6, see Figure S1.4 for identical plots for all other explanatory variables). The similarities between the distribution of the maxima of the taxon-specific responses and the community distribution demonstrates that the community distributions are sufficiently narrow to prevent overfitting without needlessly constraining the taxon-specific responses. The marginal posterior taxon-specific parameter distributions β_{kj}^{taxa} quantify the strength and uncertainty of the taxon-specific responses for each explanatory variable, and can be combined to quantify the uncertainty and relative importance of the responses for the overall community.

For instance, taxa with a narrow marginal posterior parameter β_{kj}^{taxa} distribution and a maximum close to zero have a high certainty of exhibiting no response to a given explanatory variable (Figure S1.6; see Table S1.7 for a summary of significant responses, and Appendix S1 subsection S1.10 for a table of significant responses for all taxa). The uncertainty of the overall community response to an explanatory variable can be quantified as the distribution of the relative standard deviations of the taxon-specific responses (e.g., Figure S1.7). In our analysis, we found the uncertainty of the community response was highest for livestock unit densities in the catchment, followed by flow velocity and riparian forest cover (Figure S1.7).

The relative importance of the explanatory variables (which differ in their units and the range of environmental conditions covered in the data; e.g., Figure S1.2) to community composition can be assessed by comparing the degree to which the linear predictor (Equation 1) for multiple taxa (and thus, the overall community) is shifted by each explanatory variable (e.g., Figure S1.3). Our results revealed stream temperature to have the largest effect on the community in all models (Figure S1.3).
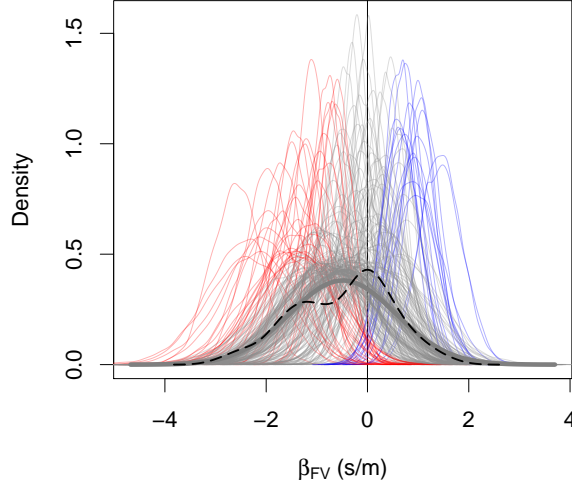
Figure 6: In the hierarchical mSDM, the marginal posterior distribution of the community response β_k^{comm} (bold grey line) constrains the taxon-specific parameters $\beta_{kj}^{\mathrm{taxa}}$ (thin transparent lines) to flow velocity, with the maxima of the taxon-specific responses (dashed line) closely following the community response. The diverse taxon-specific responses indicate significant positive (blue), negative (red), and non-significant (grey) responses based on the 5th and 95th quantile of the marginal posterior distributions of the taxon-specific parameters $\beta_{kj}^{\mathrm{taxa}}$..

In addition to quantifying taxon-specific responses and their relative uncertainty to explanatory variables, a simple mSDM (UF0) can be used to quantify the probabilities of occurrence for any individual taxon in response to varying environmental conditions (e.g., the amphipod family Gammaridae; Figure 7a). These "response curves" visualize the relative effects of the explanatory variables on the distributions of taxa, and can be compared with the geographic distribution of the probabilities of occurrence and observations of a taxon to show regions with a good (i.e., large blue and/or small red points) or bad (i.e., large red and/or small blue points) quality of fit for individual observations (e.g., Figure 7a, Figure 7b; see Appendix S1 subsection S1.10 for similar plots and maps of all taxa in the individual models and the mSDM UF0). Large differences in the probability of occurrence (i.e., point size) for an individual observation indicate a high uncertainty in the modelled output (Figure 7b).
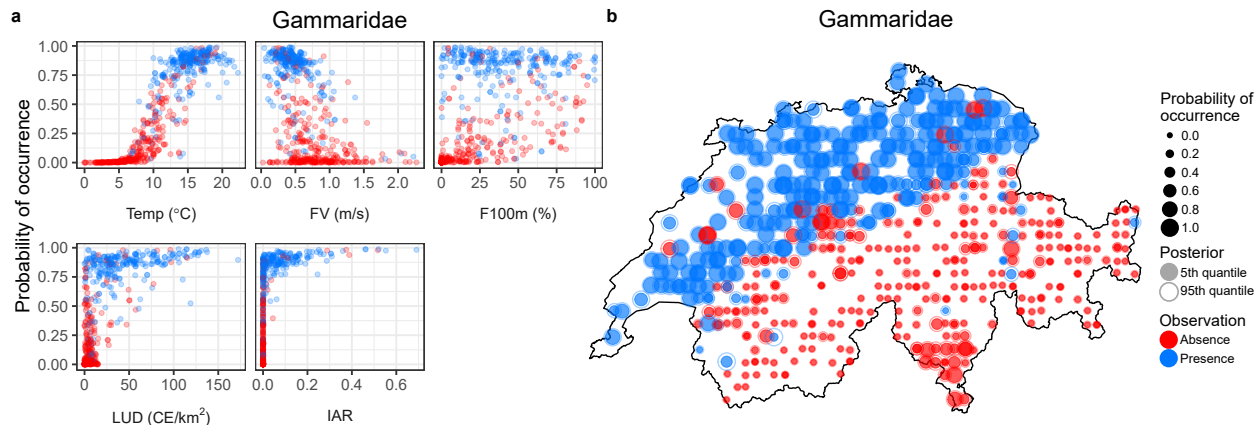
Figure 7: An example of the taxon-specific responses quantified in a hierarchical multi-species model, including (a) the calibrated probabilities of occurrence and observations versus the explanatory variables and (b) the geographic distribution of the probabilities of occurrence versus the observations of the amphipod family Gammaridae (265 occurrences, $D^2 = 0.58$) in the mSDM UF0. Observations are colored for a presence (blue) or absence (red). In (a) the probabilities of occurrences are at the maximum posterior distribution. In (b), the probability of occurrence increases with point size and the point fill corresponds to the probability of occurrence at the 5th (filled) and 95th (unfilled) quantiles of the marginal posterior distributions of the probabilities of occurrence.

4 Discussion

In comparing the performance of individual (iSDMs), hierarchical multi-species (mSDMs), and joint species distribution models (jSDMs) when applied to a high-quality dataset of stream invertebrate communities, we found that an mSDM with a relatively simple hierarchical structure (i.e., model UF0; see Table 1) achieves a higher predictive performance relative to other modelling approaches (Figure 4, Figure 5). To comprehensively demonstrate how model performance varies within a community, we applied iSDMs using the same explanatory variables to the entire community while acknowledging that in practice, iSDMs with fewer explanatory variables would have to be used for taxa with insufficient presence observations (as the underlying parameters of the full model are unidentifiable for rare taxa; see Figure S1.8). Compared to iSDMs, the simplest mSDM performs reasonably well for specific taxa when calibrated to the entire dataset (Figure 3, Figure 7). The iSDMs provide unconstrained parameter estimates and appear to show minor improvements in the quality of fit for the community relative to the simplest mSDM, with indications of overfitting for taxa with insufficient presence observations and unidentifiable parameters (Figure 3, Figure 4). The predictive performance of the individual models deteriorates further during cross-validation, providing evidence of overfitting (Figure 4).

For taxa that are neither too common nor too rare and well-explained by the explanatory variables, the simplest mSDM achieves a quality of fit similar to the iSDMs, while mitigating overfitting for rare taxa due to overarching community parameter distributions that constrain the taxon-specific parameters (Figure 4). The hierarchical model structure allows rare taxa with unidentifiable parameters to "borrow" information from the community, thereby improving predictive performance for the community and rare taxa in particular. Unlike identical iSDMs, the mSDM can be applied to a whole community without the need to exclude taxa (e.g., Calabrese et al., 2014; Kuemmerlen et al., 2015; Pottier et al., 2013). Fitting all taxa with iSDMs would require the selection of taxon-specific explanatory variables, with fewer variables for rare taxa (or simply fitting the parameters α_j^{taxa} for very rare taxa and setting all β_{kj}^{taxa} to zero). An additional advantage of the mSDM is that it combines the analytical possibilities of iSDMs and a RM by quantifying the relative effect size and its uncertainty for explanatory variables for the overall community and specific taxa (Figure 6, Figure S1.3, Figure 7), and community properties (such as family and species richness) similar to an equivalent RM (Figure 5).

An mSDM can be extended with more complex model structures that provide additional analytical possibilities and may improve model performance. For instance, we introduced correlated community parameters in the mSDM (i.e., in model CF0) with the aim of testing whether the identifiability of the taxon-specific parameters could be improved, but found no change in the overall quality of fit or predictive performance (Equation 14, Figure 4). Still, the correlations between community parameters can be informative (e.g., Figure S1.5), and reduce the effective number of parameters relative to the simplest mSDM (see subsection S1.4). The inclusion of site effects in the mSDM (e.g., as in model CT0) can facilitate the selection of additional explanatory variables to include in the model or be compared with existing variables to identify potential non-linear responses to environmental conditions (e.g., Figure S1.10, Figure S1.11). In this particular application, we included a quadratic transformation of stream temperature as an additional parameter due to the wide variation in the topography of Switzerland (Figure S1.4). This allows us to quantify preferences for intermediate temperatures in addition to montonic dependencies on temperature. However, we found that while the site effect provides an increase in the quality of fit relative to a simpler mSDM, the predictive performance declines as the site effect is unknown at sites not used during model calibration.

Extending the mSDM (with the site effect) into a joint model allowed us to assess whether the inclusion of residual correlations between taxa (e.g., Figure S1.12) improves predictive performance. The residual correlations between taxa can provide indications either of biotic interactions or joint responses to unknown explanatory variables. In our study, we found that a joint model (parameterized with latent variables) further improved the quality of fit (Figure 3), although predictive performance progressively declined with the number of latent variables included in the model (Figure 4; our trials found that the predictive performance continued to deteriorate with joint models including up to four latent variables [which we do not show here]). Similarly, community properties including family and species richness show an increase in predictive bias and uncertainty with more complex mSDMs and jSDMs, while the predicted richness of the simplest mSDM has a lower predictive uncertainty relative to more complex models (though predictions remain biased; Figure 5).

This particular result contradicts our initial hypothesis that the joint model would improve predictive performance, as similar studies have found that joint models generally improve the quality of fit and predictive performance (relative to individual models) for various communities (Clark et al., 2014; Harris, 2015; Ovaskainen et al., 2016a,1; Taylor-Rodríguez et al., 2017; Tikhonov et al., 2017). For example, Clark et al. (2014) and Taylor-Rodríguez et al. (2017) show that joint models of forest communities improve predictive performance and the prediction of species richness over equivalent individual models (Guisan and Rahbek, 2011), with a similar study of plant communities confirming these findings (e.g., Tikhonov et al., 2017). Studies applying joint models to freshwater communities suggest that shared responses to environmental conditions may be more important in determining community composition than residual correlations between taxa (Inoue et al., 2017; Pollock et al., 2014; Royan et al., 2016).

We speculate that, contrary to forest communities with biotic interactions that are dominated by competition (Clark et al., 2014; Taylor-Rodríguez et al., 2017), freshwater invertebrate communities have more diverse energy sources such as alternative food sources, multiple functional feeding types, and frequent stream disturbance events (preventing the ecosystem from reaching carrying capacity), thereby reducing the overall importance of biotic interactions for predictive modelling. We urge a careful interpretation of correlations between stream invertebrate taxa (e.g., Figure S1.12) as the effects of unknown, external explanatory variables and the type of biotic interactions (e.g., competition versus predation) cannot be distinguished (Clark et al., 2014; Dormann et al., 2018). Although the joint model did not improve predictive performance in this particular application, joint models that do not show a significant deterioration in predictive performance can be used to test for prior hypotheses of interactions between specific species, and to facilitate discussions with biologists to stimulate additional model development (Dormann et al., 2018). An investigation of the ecological context in which the inclusion of residual dependencies among distributions of taxa can lead to improved predictive performance would be an interesting topic for future research. An additional difficulty of applying a joint model to our data set is the large number of observed taxa, making an unstructured approach to estimating correlation parameters infeasible. Parameterization of the correlations with latent variables provides a method to reduce the number of parameters, however this comes at the cost of inducing structure to the correlation coefficients, making it difficult to identify a few, important correlations among the many taxa.

A notable limitation to the modelling approaches applied in this study is that the stacked iSDMs, mSDMs, and jSDMs impose an identical set of explanatory variables on the community, which may not be realistic (particularly for diverse communities with a large number of taxa). Given that iSDMs allow for the selection of taxon-specific variables, we considered statistical and machine learning techniques often used to guide variable selection for iSDMs, including stepwise selection methods (e.g., Leps et al., 2015), elastic net regularization (Zou and Hastie, 2005), and random forests (Breiman, 2001; Cutler et al., 2007). While these techniques are suitable for selecting variables specific to each taxon, we instead opted for a computationally intensive but feasible "exhaustive search" procedure that effectively combines cross-validation with a best subsets approach (see James et al., 2013). To reduce the computational burden of the procedure, we limited the number of potential individual mod-

els for cross-validation (i.e., of all taxa in the community) by excluding models with collinear explanatory variables and limiting model complexity. With this approach, we comprehensively quantified the predictive performance of all potential sets of independent explanatory variables for multiple taxa (i.e., excluding rare taxa with unidentifiable parameters) in the community rather than for only specific taxa, while avoiding the disadvantages of stepwise selection or regularization techniques (e.g., variable selection in the presence of collinearity).

5   Conclusions

In this study, we comprehensively compared the performance of increasingly complex individual, hierarchical multi-species, and joint species distribution models when applied to a diverse community of freshwater invertebrates. In extending these models, we demonstrated the analytical possibilities of additional model elements, and quantified their effect on model performance (e.g., quality of fit, predictive performance). Our findings show that some of the limitations of modelling multiple taxa in a community independently of one another (i.e., with individual models that have difficulties in identifying parameters and overfit for rare taxa when using the same explanatory variables for all models) can be overcome by constraining the taxon-specific parameters with overarching community parameters. By assuming that taxa contribute to and follow an overarching community response in a hierarchical multi-species distribution model, rare taxa with few data points can "borrow" information from the community, thereby improving the predictive performance for the community and rare taxa in particular. Including correlations between the community parameters supports the identification of interactions between taxon-specific responses to explanatory variables, but did not increase the predictive power of the model in our study. Similarly, site-effects can be used to identify additional variables to include in the model, but did not increase predictive performance in our case study. To test whether the inclusion of residual correlations between taxa improves predictive performance, we parameterized a correlation term of the hierarchical multi-species model with latent variables to obtain a joint model. In our application, the predictive performance of the joint model did not improve compared to the simpler mSDMs, indicating either that the biotic interactions or joint responses to unknown factors are less important than the findings of other studies applying joint models (e.g., for plant communities), or that the latent variable approach was not sufficient to detect relevant correlations. We found that a hierarchical multi-species model predicts richness as well as a richness model, and additionally quantifies the response of the overall community and specific taxa (with their respective uncertainties) to selected explanatory variables. Overall we view hierarchical and joint SDMs as promising tools that combine and extend many of the analytical possibilities of individual models, and urge ecologists interested in exploring community data to select a parsimonious modelling approach that is adequate in reflecting the dominant environmental conditions that drive species distributions.

Acknowledgements

Biosketches

Bogdan Caradima is interested in applying geographic information science, spatial and statistical models, and multi-criteria decision analysis to inform environmental management of freshwater and forest ecosystems, and to analyze location decisions in economic geography (http://www.eawag.ch/~caradibo).

Nele Schuwirth is interested in the development of mechanistic models for freshwater ecosystems to reveal the influence of multiple stressors on community structure and ecosystem function, ecological assessment, and multi-criteria decision support for environmental management (http://www.eawag.ch/~schuwine).

Peter Reichert is interested in combining mechanistic and statistical approaches for hydrological and ecological modelling, quantifying prediction uncertainty, and using these predictions in environmental decision support (http://www.eawag.ch/~reichert).

References

Aitken, M., Roberts, D. W., and Shultz, L. M. (2007). Modelling distributions of rare plants in the great basin, Western North America. *Western North American Naturalist*, 67(1):26–38.

Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, 200(1–2):1–19.

BDM Coordination Office (2009). Biodiversity Monitoring in Switzerland (BDM) Coordination Office 2009: The state of biodiversity in Switzerland. Overview of the findings of Biodiversity Monitoring Switzerland (BDM) as of May 2009. Technical Report 0911, Federal Office for the Environment (FOEN).

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., and White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3):127–135.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Breiner, F. T., Guisan, A., Bergamini, A., and Nobis, M. P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, 6(10):1210–1218.

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

Calabrese, J. M., Certain, G., Kraan, C., and Dormann, C. F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, 23(1):99–112.

Clark, J. S., Gelfand, A. E., Woodall, C. W., and Zhu, K. (2014). More than the sum of the parts: forest climate response from joint species distribution models. *Ecological Applications*, 24(5):990–999.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J. (2007). Random Forests for Classification in Ecology. *Ecology*, 88(11):2783–2792.

Dormann, C. F., Bobrowski, M., Dehling, D. M., Harris, D. J., Hartig, F., Lischke, H., Moretti, M. D., Pagel, J., Pinkert, S., Schleuning, M., Schmidt, S. I., Sheppard, C. S., Steinbauer, M. J., Zeuss, D., and Kraan, C. (2018). Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography*, pages 1–13.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222.

Elith, J., Catherine, G., P Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J Hijmans, R., Huettmann, F., R Leathwick, J., Lehmann, A., Li, J., et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2):129–151.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N. G., Lehmann, A., and Zimmermann, N. E. (2006). Using Niche-Based Models to Improve the Sampling of Rare Species. *Conservation Biology*, 20(2):501–511.

Guisan, A., Edwards, T. C., and Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling*, 157(2):89–100.

Guisan, A. and Rahbek, C. (2011). SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, 38(8):1433–1444.

Guisan, A. and Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2–3):147–186.

Harris, D. J. (2015). Generating realistic assemblages with a joint species distribution model. *Methods in Ecology and Evolution*, 6(4):465–473.

Hui, F. K. C., Warton, D. I., Foster, S. D., and Dunstan, P. K. (2013). To mix or not to mix: comparing the predictive performance of mixture models vs. separate species distribution models. *Ecology*, 94(9):1913–1919.

Inoue, K., Stoeckl, K., and Geist, J. (2017). Joint species models reveal the effects of

environment on community assemblage of freshwater mussels and fishes in European rivers. *Diversity and Distributions*, 23(3):284–296.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*, volume 112. Springer.

Kaelin, K. and Altermatt, F. (2016). Landscape-level predictions of diversity in river networks reveal opposing patterns for different groups of macroinvertebrates. *Aquatic Ecology*, 50(2):283–295.

Kuemmerlen, M., Schmalz, B., Cai, Q., Haase, P., Fohrer, N., and Jähnig, S. C. (2015). An attack on two fronts: predicting how changes in land use and climate affect the distribution of stream macroinvertebrates. *Freshwater Biology*, 60(7):1443–1458.

Latimer, A. M., Banerjee, S., Sang Jr, H., Mosher, E. S., and Silander Jr, J. A. (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(2):144–154.

Leps, M., Tonkin, J. D., Dahm, V., Haase, P., and Sundermann, A. (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.

Letten, A. D., Keith, D. A., Tozer, M. G., and Hui, F. K. C. (2015). Fine-scale hydrological niche differentiation through the lens of multi-species co-occurrence models. *Journal of Ecology*, 103(5):1264–1275.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9):1989–2001.

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J., and Guisan, A. (2010). Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, 143(11):2647–2657.

Mouton, A. M., De Baets, B., and Goethals, P. L. M. (2010). Ecological relevance of performance criteria for species distribution models. *Ecological Modelling*, 221(16):1995–2002.

Nelder, J. A. and Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313.

Ovaskainen, O., Abrego, N., Halme, P., and Dunson, D. (2016a). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, 7(5):549–555.

Ovaskainen, O., Hottola, J., and Siitonen, J. (2010). Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, 91(9):2514–2521.

Ovaskainen, O., Roy, D. B., Fox, R., and Anderson, B. J. (2016b). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution*, 7(4):428–436.

Ovaskainen, O. and Soininen, J. (2011). Making more out of sparse data: hierarchical modeling of species communities. *Ecology*, 92(2):289–295.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., and Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 20(5):561–576.

Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A., and McCarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, 5(5):397–406.

Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C. F., Vittoz, P., and Guisan, A. (2013). The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology and Biogeography*, 22(1):52–63.

R Core Team (2018). R: A Language and Environment for Statistical Computing.

Royan, A., Reynolds, S. J., Hannah, D. M., Prudhomme, C., Noble, D. G., and Sadler, J. P. (2016). Shared environmental responses drive co-occurrence patterns in river bird communities. *Ecography*, 39(8):733–742.

Stan Development Team (2016). *RStan: the R interface to Stan*.

Stucki, P. (2010). Methoden zur Untersuchung und Beurteilung der Fliessgewässer. Technical Report 1026: 61S, Bundesamt für Umwelt, Bern, Bern, Switzerland.

Taylor-Rodríguez, D., Kaufeld, K., Schliep, E. M., Clark, J. S., and Gelfand, A. E. (2017). Joint Species Distribution Modeling: Dimension Reduction Using Dirichlet Processes. *Bayesian Analysis*, 12(4):939–967.

Thorson, J. T., Ianelli, J. N., Larsen, E. A., Ries, L., Scheuerell, M. D., Szuwalski, C., and Zipkin, E. F. (2016). Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring: Joint dynamic species distribution models. *Global Ecology and Biogeography*, 25(9):1144–1158.

Tikhonov, G., Abrego, N., Dunson, D., and Ovaskainen, O. (2017). Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context. *Methods in Ecology and Evolution*, 8(4):443–452.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition.

Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., and Hui, F. K. C. (2015a). So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology & Evolution*, 30(12):766–779.

Warton, D. I., Foster, S. D., De'ath, G., Stoklosa, J., and Dunstan, P. K. (2015b). Model-based thinking for community ecology. *Plant Ecology*, 216(5):669–682.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.