



Canadian Journal of Fisheries and Aquatic Sciences

Improved understanding and prediction of freshwater fish communities through the use of joint species distribution models

Journal:	<i>Canadian Journal of Fisheries and Aquatic Sciences</i>
Manuscript ID	cjfas-2019-0348.R2
Manuscript Type:	Article
Date Submitted by the Author:	11-May-2020
Complete List of Authors:	Wagner, Tyler; US Geological Survey Hansen, Gretchen J. A.; Univ Minnesota Schliep, Erin; University of Missouri System, Statistics Bethke, Bethany J.; Minnesota Dept Nat Resources, Fisheries Research Honsey, Andrew; University of Minnesota, Ecology, Evolution, and Behavior Jacobson, Peter; Minnesota Department of Natural Resources, Kline, Benjamin ; Pennsylvania State University, Ecosystem Science and Management White, Shannon; Pennsylvania State University,
Keyword:	joint species distribution models, fish communities, species interactions
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

SCHOLARONE™
Manuscripts

1 Improved understanding and prediction of freshwater fish communities through the use of joint
2 species distribution models

3

4 Tyler Wagner¹, Gretchen J.A. Hansen², Erin M. Schliep³, Bethany J. Bethke⁴, Andrew E.
5 Honsey², Peter C. Jacobson⁴, Benjamin C. Kline⁵, Shannon L. White⁵

6

7 ¹U.S. Geological Survey, Pennsylvania Cooperative Fish and Wildlife Research Unit,
8 Pennsylvania State University, University Park, Pennsylvania, USA

9 ²Department of Fisheries, Wildlife, and Conservation Biology, University of Minnesota, St. Paul,
10 MN 55188, USA

11 ³Department of Statistics, University of Missouri, 146 Middlebush Hall,
12 Columbia, Missouri 65211, USA

13 ⁴Minnesota Department of Natural Resources, Division of Fish and Wildlife, St. Paul, MN,
14 55155, USA

15 ⁵Pennsylvania Cooperative Fish and Wildlife Research Unit, Pennsylvania State University,
16 University Park, Pennsylvania, USA

18 Abstract

19 Two primary goals in fisheries research are to (1) understand how habitat and environmental
20 conditions influence the distribution of fishes across the landscape, and (2) make predictions
21 about how fish communities will respond to environmental and anthropogenic change. In inland,
22 freshwater ecosystems, quantitative approaches traditionally used to accomplish these goals
23 largely ignore the effects of species interactions (competition, predation, mutualism) on shaping
24 community structure, potentially leading to erroneous conclusions regarding habitat associations
25 and unrealistic predictions about species distributions. Using two contrasting case studies, we
26 highlight how joint species distribution models (JSDMs) can address the aforementioned
27 deficiencies by simultaneously quantifying the effects of abiotic habitat variables and species
28 dependencies. In particular, we show that conditional predictions of species occurrence from
29 JSDMs can better predict species presence/absence compared to predictions that ignore species
30 dependencies. JSDMs also allow for the estimation of site-specific probabilities of species co-
31 occurrence, which can be informative for generating hypotheses about species interactions.
32 JSDMs provide a flexible framework that can be used to address a variety of questions in
33 fisheries science and management.

34

35 Keywords: joint species distribution models, fish communities, species interactions, multi-
36 species, species distribution models, freshwater

37 **Introduction**

38 A primary goal in fisheries science is to predict how species distributions and abundances
39 respond to current and future habitat conditions, and to use those predictions to inform fisheries
40 management (Chu et al. 2005; Xenopoulos et al. 2005; Bond et al. 2011; DeWeber and Wagner
41 2018). The field is replete with quantitative methods available for modeling the influence of
42 abiotic environmental predictors on species distributions. For example, generalized linear models
43 (GLMs) are frequently used to understand how habitat influences individual fish species
44 occurrence or abundance, the number of species present, and/or indices of biotic integrity (Wang
45 et al. 2001; Sutherland et al. 2002; Roy et al. 2007). Alternatively, multivariate ordination
46 methods, such as canonical correspondence analysis, can be used to understand fish assemblage
47 structure in different environments (Van Zuiden and Sharma 2016; Stewart et al. 2016).
48 However, neither GLMs nor multivariate methods explicitly account for potential dependencies
49 among species in the community, which is a fundamental oversight at the core of community
50 ecology. That is, environmental filtering selects a subset of species with specific phenotypes to
51 occupy a habitat (Poff 1997), but the presence and abundance of a species also depends on
52 species interactions including competition, predation, and mutualism (Pendleton et al. 2012;
53 Peoples et al. 2015). A failure to account for biotic interactions can provide an overly simplistic
54 view of community dynamics, which can hinder our understanding of assembly processes and
55 bias predictions of community structure under environmental change (Latimer et al. 2009; Clark
56 et al. 2014a; Harris 2015).

57 To date, attempts to incorporate species interactions into quantitative frameworks in
58 freshwater fisheries have largely focused on isolating the effects of two or three species in the
59 community. For example, two-species occupancy models have been used to elucidate how the

60 presence of one species modulates the response of another species to watershed stressors
61 (Wagner et al. 2013), and the presence of one species has been used to predict the occurrence of
62 another focal species (Van Zuiden and Sharma 2016; Taylor et al. 2018). In addition to these
63 analyses being limited to a subset of species in the community, they can only account for
64 asymmetric species interactions and some require *a priori* specification of dominant and
65 subordinate species.

66 By simultaneously quantifying effects of the abiotic environment and biotic interactions,
67 joint species distribution models (JSDMs) provide a framework for addressing the
68 aforementioned shortcomings of conventional fish community models. JSDMs can be thought of
69 as a special case of more general multivariate models, and have received increased attention over
70 the past decade (Norberg et al. 2019) in various subdisciplines of ecology (*e.g.*, marine fisheries,
71 Hill et al. 2017, Thorson 2019; forest tree communities, Clark et al. 2014b; and bird
72 communities, Stjernman et al. 2019), but remain underused in inland freshwater fisheries (but see
73 Inoue et al. 2017 and Radinger et al. 2019 for examples). Because of their flexibility, JSDMs
74 may prove useful for a variety of fisheries applications, including quantifying (i) heterogenous
75 data types, including occurrence and/or abundance, (ii) environmental drivers of species
76 occurrence/abundance, and (iii) correlations (dependencies) between species in the community
77 across space and time (Clark et al. 2014b; Pollock et al. 2014; Ovaskainen et al. 2017; Clark et
78 al. 2017).

79 JSDMs are described in detail in Pollock et al. (2014) and Clark et al. (2014b). Briefly, a
80 JSDM relies on a species covariance matrix to account for possible dependencies among species
81 (*i.e.*, potential biotic interactions) at a given location that are not accounted for by other fixed
82 effects (environmental predictors) in the model. Residual correlations can then be used to

83 generate hypotheses about the potential role of biotic interactions in structuring fish
84 communities. Additional analyses using JSMD results can provide further clarity about the
85 strength of co-occurrence patterns. For example, an odds ratio of occurrence among species pairs
86 can be an intuitive metric to understand how co-occurrence varies spatially among habitat types
87 (Gelfand and Shirota, in review).

88 JSMDs are particularly useful for analyses of fish community data, as fisheries data are
89 often collected across space and over time and thereby may exhibit spatial and/or temporal
90 autocorrelation. When spatial and/or temporal autocorrelation is present and not accounted for,
91 model inference may be biased and lead to increased type-I error rates. If strong dependence
92 exists, the inclusion of random effects can also improve predictions at unsampled locations in
93 time and space. For example, Godefroid et al. (2019) used Vector Autoregressive Spatio-
94 temporal (VAST) models (Thorson et al. 2015) to understand the spatio-temporal dynamics of
95 Pacific Herring off the west coast of Vancouver Island, Canada, and Thorson et al. (2016)
96 developed a joint dynamic species distribution model that accounted for spatial dependence and
97 temporal variation in species distributions to investigate community dynamics in a marine fish
98 community.

99 The development and application of complex spatio-temporal JSMDs in fisheries has
100 largely occurred in marine ecosystems, where long-term monitoring programs of economically
101 important fish stocks have produced information-rich datasets (large amounts of data over space
102 and time for many species). This data structure is often lacking in freshwater fisheries, where
103 sampling effort may be limited, surveys may be infrequent and non-random over space and time,
104 and assessments may include fishes for which we have limited information on behavior and life
105 history. Here, we illustrate the utility of JSMDs for modeling inland freshwater fish community

106 data using two case studies. Our aim is not to focus on the details and potential implications of
107 the results from each case study. Rather, we highlight broad themes associated with the use and
108 interpretation of JSDMs that will be applicable for answering a wide range of questions across a
109 diversity of fish communities. Specifically, we demonstrate how JSDMs can provide improved
110 understanding of the abiotic and biotic processes structuring fish communities, and how
111 accounting for species dependencies can improve model predictive performance. We provide
112 two contrasting case-studies, and we further develop our second case study to illustrate how to
113 better understand species co-occurrences and to investigate potential spatial autocorrelation.

114 Our case studies – Appalachian streams and north temperate lakes – represent two
115 contrasting aquatic ecosystems and fish communities. However, the research goals in both case
116 studies were the same: (1) improve our understanding of environmental filtering by quantifying
117 the variation among species in their responses to abiotic drivers and their shared responses to the
118 environment; (2) estimate residual correlations which can be used, along with information about
119 species traits, phylogeny, trophic ecology, and other information, to generate hypotheses about
120 the potential roles of biotic interactions in structuring these communities; (3) quantify spatial
121 variability in co-occurrences between species pairs; and (4) compare predictions of occurrence
122 when accounting for species dependencies to those that ignore species dependencies.

123 **Case studies**

124 *Appalachian stream fish communities*

125 Fish community data were collected from 60 wadeable streams in southwestern Virginia
126 and southern West Virginia, USA by White and Orth (2013). At most streams, fish were sampled
127 at six 50-m sites using single-pass backpack electrofishing. Species enumerations were converted
128 to site-specific presence/absence data, and only species that were found at 10 or more sites were

129 included in this analysis, which resulted in 16 species. Due to difficulty in field identification and
130 concerns about hybridization, sculpin were only identified to genus (*i.e.*, *Cottus* sp.).

131 Site-level microhabitat data were collected at transects spaced every 10 m and included
132 stream width, stream depth, and substrate size. The number of pools and size of every pool in
133 each site was also measured. Stream-level macrohabitat was quantified in Geographic
134 Information System (GIS), and included the proportion of the watershed with developed,
135 agriculture, and forested land-use as defined by the 2011 National Land Cover Dataset (NLCD).
136 Minority land-use classes including barren rock, open water, and wetland were not considered as
137 they accounted for less than 5% of land-use across all sites. The proportion of forested land-use
138 was homogenous across streams and developed and agricultural land-use were collinear.
139 Therefore, we only included the proportion of developed land in the watershed as a land-use
140 covariate.

141 *North temperate lake fish communities*

142 Fish community data were collected from lakes in Minnesota, USA from 1987 to 2016
143 via the Minnesota Department of Natural Resources (MNDNR) standard sampling program
144 (MNDNR 2017). This program employs trapnets and gillnets to sample the nearshore and deeper
145 littoral communities, respectively, at index stations throughout lakes. Sampling was repeated on
146 lakes on a 1- to 10-year rotation, depending on individual lake management needs. These surveys
147 were primarily designed to target sport fish. As such, many smaller-bodied fishes were not
148 captured in these gears. Lakes were sampled from June through September, and sampling at each
149 lake occurred at approximately the same day of the year across years. The number of nets set at
150 each lake depended on lake size (see MNDNR 2017 for methodological details).

151 We only considered surveys in which at least three nets of each gear type (gillnets and
152 trap nets) were deployed, and an individual lake was only included in the analysis if it was
153 sampled at least twice from 1987 to 2016. Data across years were combined to derive a single
154 presence/absence metric for each species for each lake, and non-fish species (*e.g.*, turtles and
155 crayfish) and fish not identified to species (including hybrids) were removed. Because standard
156 fisheries surveys target littoral species, supplemental data from deeper, non-standard nets and
157 angler surveys were added for coldwater fish (including lake trout *Salvelinus namaycush*, lake
158 whitefish *Coregonus clupeaformis*, and cisco *Coregonus artedi*) that inhabit deeper portions of
159 Minnesota lakes. Finally, we removed species that were not present in at least 5% of lakes and
160 lakes for which the full suite of environmental predictor variables were unavailable (see below),
161 resulting in a total of 1,213 lakes and 29 species included in the final analysis.

162 Our treatment of the fish community data (*i.e.*, removing rare species, requiring that both
163 gear types be used in each survey) and the use of standardized sampling was done, in part, to
164 reduce the potential effects of environmental conditions on detection probabilities, which can
165 vary among species and gear types (Gwinn et al. 2016 and references therein). We recognize,
166 however, that imperfect detection could affect inferences, especially for those species that are
167 difficult to detect using the gear used in the case studies. The inability of biological monitoring
168 programs to detect all species in a community (*i.e.*, imperfect detection) – and its effect on
169 parameter estimation and inference – has received considerable attention in the ecological and
170 fisheries literature (*e.g.*, MacKenzie et al 2002; Kéry 2011; Gibson-Reinemer 2016; White et al.
171 2020). Some statistical modeling frameworks for fitting JSDMs can accommodate detection
172 probabilities/sampling error (*e.g.*, the R package *gjam*; Clark et al. 2017; see Statistical analysis
173 below), however, they rely on additional species-specific information (*e.g.*, sampling effort as a

174 proxy for detection). Until the work by Tobler et al. (2019), there lacked a general analytical
175 framework for linking JSDMs that quantify residual correlations and imperfect detection in an
176 occupancy modeling framework. The application of such models in freshwater fish community
177 research is challenging because these models require data sampled across generally hundreds of
178 sites and some require repeat visits over a relatively short time period. The repeat visits are
179 necessary to estimate detection probabilities and sampling over a relatively short timeframe is
180 needed to meet the assumption of closure and that the true presence/absence of a species at a site
181 does not change over time (Tobler et al. 2019). Depending on the system, sampling a large
182 number of sites and/or performing repeat visits to a subset of sites may not be logistically
183 feasible during fisheries research efforts. Thus, methods that account for imperfect detection
184 under various data collection strategies of freshwater fish communities provides a fruitful area of
185 future research.

186 Environmental variables for the lake analysis were compiled from multiple sources. Lake
187 area and maximum depth were obtained from MNDNR. We used the 2011 NLCD to calculate
188 watershed disturbance, defined as the proportion of the entire upstream watershed in agricultural
189 or urban land use (Jacobson et al. 2016). Water clarity was indexed using Secchi depth collated
190 from state agency databases (MNDNR and Minnesota Pollution control agency;
191 <https://www.pca.state.mn.us/water/water-quality-data>), and we calculated lake-level median
192 annual Secchi depth from data collected from June 1 through August 31 from 1987 to 2018. We
193 also calculated the proportion of each lake in the littoral zone, as defined by the area of lake
194 bottom in which sub-surface light levels were at least 1% of surface light levels. Daily Secchi
195 depth estimates were converted to daily light extinction coefficients (K_d) by dividing a constant
196 coefficient of 1.7 by Secchi depth (Poole and Atkins 1929) and euphotic depth was calculated as

197 $-\log(0.01)/K_d$. Lake area at or above this depth was calculated using hypsographic curves
 198 generated from digitized bathymetric maps. Lake temperature was indexed by mean annual
 199 degree days (base temperature 5°C; Chezik et al. 2014) calculated from simulated water
 200 temperatures for 1980-2015 (Winslow et al. 2017).

201 **Statistical analysis**

202 Several statistical modeling frameworks exist for fitting JSDMs (see Norberg et al. 2019
 203 for a summary of frameworks), of which we used the Generalized Joint Attribute Modelling
 204 (gjam) package in R version 3.6.1 (Clark et al. 2017; R Core Team 2019). Though both of our
 205 case studies focused on modeling fish occurrence, gjam can accommodate a variety of response
 206 data types (*e.g.*, discrete and continuous abundance, ordinal counts), including combinations of
 207 response types. However, it is worth noting that gjam cannot explicitly address temporal and
 208 spatial autocorrelations at this time (but see below for methods to assess potential spatial
 209 autocorrelation).

210 Presence/absence of each species at each stream site or lake was modeled using a
 211 multivariate probit model and latent variable parameterization (Chib and Greenberg 1998).
 212 Following the notation of Wilkinson et al. (2019), we let $y_{ij}=1$ denote presence and $y_{ij}=0$ denote
 213 absence of species j at site i , where $j = 1 \dots J$ and $i = 1 \dots n$. The basic model is as follows:

$$214 \quad y_{ij} = 1(z_{ij} > 0) \quad (1)$$

$$215 \quad z_{ij} = \mathbf{u}_{ij} + \mathbf{e}_{ij}$$

$$216 \quad \mathbf{u}_{ij} = \mathbf{X}_i \cdot \mathbf{B}_{\cdot j}$$

$$217 \quad \mathbf{e}_i \sim \text{MVN}(\mathbf{0}, \mathbf{R})$$

218 Presence and absence of each species is specified through the latent variable z_{ij} , where $z_{ij} > 0$
219 denotes presence and $z_{ij} \leq 0$ denotes absence. The latent variable is equal to the sum of the linear
220 predictor u_{ij} and correlated residual error e_{ij} .

221 Following the generalized linear model framework, the linear predictor is the product of k
222 = 1...K environmental covariates $\mathbf{X}_{i,\cdot}$ and regression coefficients $\mathbf{B}_{\cdot,j}$. Residual correlations are
223 contained in the $J \times J$ covariance matrix \mathbf{R} , which captures the dependencies in co-occurrence
224 among species after accounting for covariates in the model. That is, \mathbf{R} identifies those species
225 that co-occur more or less frequently than predicted given the environmental covariates. Residual
226 dependencies also allow for predicting the occurrence of each species conditionally on the
227 occurrence of other species in the community. For the Appalachian stream fish case study, a
228 normally distributed random stream effect (random intercept) was also included in the model to
229 accommodate multiple observations collected in each stream.

230 For both analyses, all covariates were standardized to be mean zero with standard
231 deviation one, and default priors specified in the `gjam` package were used for all parameters. For
232 the Appalachian stream fish analysis, the posterior distribution was summarized using 120,000
233 MCMC samples, sampled using a Gibbs sampler, after discarding the first 90,000 samples as
234 burn-in. For the north temperate lakes case study, 90,000 samples were used to summarize the
235 posterior distribution, after discarding the first 70,000 samples as burn-in. We calculated the
236 correlation among species pairs due to shared environmental responses following the methods of
237 Pollock et al. (2014).

238 *Spatial autocorrelation*

239 Potential residual spatial autocorrelation was examined by creating spline correlograms
240 for each species using the `ncf` package in R (Bjørnstad 2020). Spline correlograms display the

241 spatial correlation as a function of distance that is smoothed using a spline function (Bjørnstad
 242 and Falck 2001). Five hundred bootstrap iterations were used to obtain 95% pointwise
 243 confidence intervals. Correlation estimates and confidence intervals that deviate from zero
 244 indicate the presence of spatial autocorrelation. (Bjørnstad and Falck 2001). Although we use
 245 Euclidean distance to assess the presence of spatial autocorrelation for the lakes case study,
 246 stream-network distance could be considered for accommodating spatial autocorrelation in lotic
 247 systems (ver Hoef 2006; Hocking et al. 2018).

248 *Odds ratios and species co-occurrences*

249 While residual correlations among species pairs are often used to generate hypotheses
 250 about potential species interactions and co-occurrence patterns (*sensu* Pollock et al. 2014;
 251 Radinger et al. 2018; Tobler et al. 2019), they provide little interpretable information about
 252 whether species pairs are likely to co-occur (positive dependency) or not (negative dependency)
 253 at a given location. Therefore, we also employed the methods of Gelfand and Shirota (in review)
 254 to calculate odds ratios between species pairs to obtain interpretable measures of species
 255 dependencies across locations. Importantly, these values take into account the response of each
 256 species to environmental conditions at each location in addition to the residual dependence
 257 between species. We calculated the odds ratio ($\theta_i^{j,j'}$) for species j and j' at site i as:

$$258 \theta_i^{j,j'} = (P(Z_{ij} < 0, Z_{ij'} < 0)P(Z_{ij} \geq 0, Z_{ij'} \geq 0)) / (P(Z_{ij} \geq 0, Z_{ij'} < 0)P(Z_{ij} < 0, Z_{ij'} \geq 0)) \quad (2)$$

259 The joint probabilities in the numerator represent the probability that species j and j' are absent at
 260 site i and the probability that species j and j' co-occur at site i , respectively. The joint
 261 probabilities in the denominator represent the probability that species j is present and species j' is
 262 absent at site i and the probability that species j is absent and species j' is present at site i ,
 263 respectively. Put another way, the numerator represents joint occurrence and joint absence

264 (sympatry), while the denominator measures the probability that each species occurs separately
265 (allopatry). As such, if there is positive dependence between species the odds ratio will be > 1 , if
266 there is negative dependence the odds ratio will be < 1 , and the odds ratio will equal 1 for the
267 case of species independence. For site i , we report $\log_{10}(\theta_i^{j,j'})$, where a value of zero indicates that
268 presence/absence of species j and j' are pairwise independent, values > 0 indicate positive
269 pairwise dependence, and values < 0 indicate negative pairwise dependence.

270 *Model performance and predictive comparisons*

271 When correlations between pairs of species are large, conditional predictions can vary
272 substantially from unconditional (marginal) predictions (Tikhonov et al. 2017). Thus, for both
273 case studies we compared predictive performance by calculating the marginal and conditional
274 posterior mean area under the estimated receiver operating characteristic curve (AUC) values for
275 each species. We considered models with AUC values < 0.7 to be low accuracy and of limited
276 use, models with AUC values > 0.7 but < 0.9 to be of moderate accuracy, and models with AUC
277 values > 0.9 to be highly accurate most useful for interpretation and prediction (Manel et al.
278 2001). Marginal probability of occurrence predictions were obtained for each species at each
279 site/lake and were equivalent to predicting the occurrence of each species independently, thus
280 ignoring potential species dependencies. Conditional probability of occurrence predictions were
281 obtained for each species and each site/lake given the presence/absence of all other species at the
282 given location. In each case we calculated the posterior mean AUC and associated 95% credible
283 intervals.

284 To gain additional insight into model predictive performance, marginal and conditional
285 sensitivity (the ability of the model to correctly predict species presence) and specificity (ability
286 of the model to correctly predict species absence) were also calculated. Calculating sensitivity

287 and specificity requires a threshold probability separating predicted presences from predicted
288 absences, and we used a threshold that minimized overall misclassification error implemented
289 using the InformationValue package in R (Prabhakaran 2016). For illustrative purposes, we only
290 report results of sensitivity and specificity for the north temperate lakes case study.

291 **Results and discussion**

292 *Appalachian stream fish communities*

293 Effects of abiotic environment

294 Overall, environmental covariates were generally poor predictors of stream fish
295 occurrence, and significant effects differed in direction and magnitude (Figure 1). For example,
296 mean substrate size was an important predictor for three species; two were more likely to occur
297 at sites with smaller substrate (white sucker *Catostomus commersonii* and rock bass *Ambloplites*
298 *rupestris*), and redbreast sunfish *Lepomis auritus* was more likely to occur at sites with larger
299 substrates. Two species (white sucker and bluegill *Lepomis macrochirus*) were more likely to
300 occur at wider stream sites, and two species (white shiner *Luxilus albeolus* and fantail darter
301 *Etheostoma flabellare*) were more likely to occur in developed watersheds.

302 Residual and shared environmental correlations

303 Correlations between species can be due to a shared response to environmental conditions
304 and/or pairwise residual correlations that capture potential biotic interaction. Shared
305 environmental correlations suggest that species pairs covary (either negatively or positively) due
306 to their response to the same abiotic conditions. Given that few habitat covariates were important
307 for predicting stream fish occurrence, it was not surprising that we identified only one significant
308 pairwise shared environmental correlation: rock bass and white sucker were positively correlated
309 due to their shared environmental response (Figure 2).

310 After controlling for the effects of environmental covariates, the residual correlations for
311 several species pairs were significant. Most of these correlations were positive, but two were
312 negative (Figure 2). The two negative residual correlations were between *Cottus* spp. and creek
313 chub *Semotilus atromaculatus* and *Cottus* spp. and Clinch dace *Chrosomus* sp. cf. *saylori*. Due to
314 the fact that very few environmental covariates were significant for predicting species
315 occurrence, it is likely that many of these residual correlations are due to missing habitat
316 covariates in the model rather than representing potential biotic interactions. However, it is
317 possible that the positive residual correlation between creek chub and Clinch dace is indicative of
318 a mutualistic relationship in the form of a nest association (White and Orth 2014; but see below
319 for cautions with this interpretation).

320 Predictive performance

321 As expected, given that abiotic habitat variables poorly predicted species occurrence,
322 marginal predictive performance measured by AUC was poor, with most posterior mean AUC
323 values near 0.5 (no better than random; Mantel et al. 2001) for all species (Figure 3). Predictive
324 performance improved significantly for most species when conditioning predictions on the
325 occurrence of other species at a site; for most species, conditional predictions resulted in models
326 classified as moderate or high accuracy based on their AUC values (blue points in Figure 3).
327 Overall, this case study highlights that predictive models of species occurrence that are informed
328 by the presence or absence of other species in the community at a given site (*i.e.*, accounting for
329 their residual correlations among species) can be useful for predicting species occurrence, even
330 when available abiotic habitat variables are relatively uninformative.

331 *North temperate lake fish communities*

332 Effects of abiotic environment

333 In contrast to the Appalachian stream fish community analysis, abiotic habitat covariates
334 significantly predicted occurrence for many north temperate lake fishes (Figure 4). The
335 probability of presence for nearly all species increased with both lake area and maximum depth.
336 For other covariates, the direction and magnitude of effects varied among species. For example,
337 mean degree days positively influenced the presence of warmwater species like white crappie
338 *Pomoxis annularis* and common carp *Cyprinus carpio*, and negatively affected the presence of
339 coldwater species such as burbot *Lota lota* and cisco *Coregonus artedii*. The effect of water
340 clarity also varied substantially among species, with some species such as white crappie found
341 more frequently in turbid sites with low Secchi depths, and other species including largemouth
342 bass *Micropterus salmoides* more likely to occur in clear lakes with high Secchi depths.

343 Residual and shared environmental correlations

344 In lake fish communities, significant shared environmental correlations existed between
345 many species pairs (Supplemental Figure S1 and Figure 5). Given that many species were well
346 predicted by habitat variables in this analysis, positive environmental correlations indicate that
347 species prefer similar habitat conditions. For example, green sunfish *Lepomis cyanellus* were
348 significantly positively correlated with a number of species such as largemouth bass, bluegill
349 *Lepomis macrochirus*, and pumpkinseed *Lepomis gibbosus*, due to the species' similar
350 preference for warm lakes with a proportionally large littoral zone. Conversely, green sunfish
351 were negatively correlated with the presence of several species that prefer cooler lakes with
352 proportionally smaller littoral zones, including walleye *Sander vitreus*, burbot, and smallmouth
353 bass *Micropterus dolomieu* (Figures 4 and 5).

354 After controlling for the effects of environmental covariates, several significant pairwise
355 species residual correlations were identified, most of which were positive (Figure 5). For

356 example, walleye and yellow perch *Perca flavescens* exhibited positive residual correlation,
357 suggesting that these species co-occur more often than predicted given the abiotic habitat
358 covariates. In contrast, negative residual correlations between species, such as between
359 smallmouth bass and northern pike *Esox lucius*, suggest that these species co-occur less
360 frequently than expected given abiotic conditions. Such negative species correlations indicate the
361 possibility of antagonistic species interactions, where the presence of one species reduces the
362 probability of presence of another.

363 For many species pairs, both shared environmental and residual correlations were
364 significant, albeit not always in the same direction. For example, largemouth bass and bluegill
365 were positively correlated for both metrics, suggesting that these species are highly likely to co-
366 occur based on both their environmental preferences and other unmeasured factors, which may
367 include biotic relationships. In contrast, largemouth bass and walleye were negatively correlated
368 in environmental preferences, but exhibited a positive residual correlation, perhaps due to
369 widespread walleye stocking into lakes where environmental conditions are not optimal for
370 walleye.

371 Spatial autocorrelation

372 Residual spatial autocorrelation varied by species. For most species, the magnitude of
373 spatial autocorrelation was generally low, with the lower bound of the 95% confidence interval
374 overlapping or near zero (Figure S2). However, the residual spatial autocorrelation for a few
375 species (*e.g.*, bowfin *Amia calva*, greater redhorse *Moxostoma valenciennesi*) was moderate in
376 magnitude. Given that spatial autocorrelation was minimal in our data, we did not include a
377 random effect in our model to account for the correlation structure. However, in situations where
378 spatial autocorrelation is stronger, spatially varying random effects could be included to account

379 for this residual autocorrelation to help reduce bias and quantify uncertainty in the parameter
380 estimates. These random effects may also aid in the prediction of species co-occurrence at
381 unobserved locations.

382 Lake-specific odds ratios

383 For illustrative purposes, we present the log odds ratios for three species pairs: bluegill
384 and largemouth bass; burbot and golden shiner *Notemigonus crysoleucas*; and smallmouth bass
385 and walleye. Bluegill and largemouth bass commonly co-occur in north temperate lakes (Treibitz
386 et al. 1997), so we expected positive log odds ratios for most lakes for this species pair. In
387 contrast, burbot are a coldwater fish that occupies deep, clear lakes, whereas golden shiner are
388 primarily littoral and can tolerate warm, turbid waters (Figure 4). In this case, we predicted that
389 most lakes would have negative log odds ratios because burbot and golden shiner would not
390 often co-occur. Smallmouth bass and walleye can occupy similar habitats, but we had no *a priori*
391 expectation as to co-occurrence patterns in our study lakes.

392 As expected, log odds ratios were generally positive for bluegill and largemouth bass,
393 with most lakes showing strong positive dependencies (relatively large positive value of θ ; only
394 2% (24 lakes) had a negative log odds ratio). In contrast, burbot and golden shiner demonstrated
395 negative dependencies in most lakes, with only 7% (86 lakes) having a positive log odds ratio.
396 The log odds ratios for smallmouth bass and walleye varied substantially, with many lakes
397 displaying positive and negative dependencies (Figure 6). In addition, the log odds ratio for the
398 majority of the lakes was near zero for these two species, indicating pairwise independence in
399 occurrence (*i.e.*, no relationship between the occurrence of one species and the other). In the
400 future, research focused on elucidating potential lake and landscape drivers of the spatial

401 variability in dependencies between species may improve our understanding of important
402 determinants of co-occurrence patterns.

403 Predictive performance

404 The increased number of significant abiotic habitat covariates in the north temperate
405 lakes analysis relative to the stream analysis led to improved marginal predictions (*i.e.*, AUC was
406 generally greater than 0.7 for marginal predications for several species; Figure 7). Including
407 species dependencies increased predictive accuracy for most, but not all, species (Figure 7).
408 Model sensitivity and specificity varied along a gradient of species prevalence (Figure 8).
409 Specifically, sensitivity was generally high for common species and specificity was high for rare
410 species, indicating that the model generally predicted common species to be present in, and rare
411 species to be absent from, most lakes.

412 Importantly, conditional predictions still often significantly outperformed marginal
413 predictions with respect to predicting where common species are absent and where rare species
414 are present (Figure 8). This suggests that, even when environmental variables are useful for
415 predicting fish species occurrence, leveraging information on the fish community can still
416 improve predictive accuracy in identifying where species are most likely to occur. It should be
417 mentioned, however, that the sensitivity and specificity values and relationship with species'
418 prevalence that we report are a function of the probability cutoff that we used for separating
419 predicted presences from absences (*i.e.*, using probabilities that minimized overall
420 misclassification error). When alternative criteria were used (in this case, minimizing sensitivity
421 + specificity), the relationship between prevalence and model accuracy disappeared, although
422 overall accuracy was lower on average (Figure S3). Probability thresholds can be selected based

423 on the relative importance of predicting true presences vs. true absences, and this decision should
424 be made on a case-by-case basis to reflect management or research priorities.

425 **Summary and cautions**

426 *Inferring species interactions from co-occurrence*

427 JSDMs account for potential dependencies among species in a fish community. However,
428 the estimated associations are correlational and are influenced by environmental predictors –
429 including both those used in the model and important environmental variables that are left out.
430 Residual correlations may also be a function of not only biotic interactions, but also phylogenetic
431 history, biogeography, stocking, and shared migration history (Dormann et al. 2018).
432 Accordingly, and as with other methods used to infer species interactions (Dormann et al. 2018),
433 species associations derived from JSDMs should be interpreted with caution. Specifically,
434 because of their correlational nature and the fact that species associations are influenced by many
435 factors, estimated species associations are best viewed as a tool for generating hypotheses, not
436 for inferring direct species interactions that lead to changes in fitness, distribution or abundance.
437 In fact, Barner et al. (2018) compared estimated species associations from a variety of co-
438 occurrence models to experimentally determine net and direct species interactions and found that
439 all methods performed poorly at recovering empirical species interactions. Zurell et al. (2018)
440 also found through a simulation exercise that the residual correlations from JSDMs performed
441 poorly at capturing known interaction effects. However, other approaches applied to spatio-
442 temporal datasets have been used to help elucidate species interactions (Thorson et al. 2017).

443 The inability to identify species interactions from species associations could be due to (1)
444 scale dependency - the strength and sign of inferred interactions can depend upon the scale of
445 observation; (2) complex biotic interactions - there can be several types of interactions for a

446 given species pair and these can be context dependent; and/or (3) circularity - species
447 interactions can drive co-occurrence, but co-occurrences are themselves necessary for
448 interactions (Thurman et al. 2019). In addition, the ability to infer species interactions from
449 associations may be dampened by the presence of multiple trophic levels in the community.
450 Specifically, Thurman et al. (2019) demonstrated that including a predator into a community
451 composed solely of competitors reduced the ability to infer competitive interactions. Dormann et
452 al. (2018) provide a useful set of questions related to data and models that can be evaluated in an
453 effort to determine the strength of inferences that might be made for a given study attempting to
454 infer species interactions from co-occurrence data.

455 If understanding species interactions or generating hypotheses about potential
456 interactions from co-occurrence data is a primary research goal, then methods that account for
457 both indirect and direct species interactions may perform better than models that only estimate
458 net effects, such as residual correlations from JSDMs. In fact, Thurman et al. (2019) found that a
459 Markov network model (Harris 2016) that estimated direct effects between species while
460 controlling for indirect effects among all species in the community outperformed several other
461 methods, including residual correlations from a JSDM, at inferring known species interactions.
462 There are several existing methods that allow for estimating direct interactions among species
463 pairs while accounting for indirect interactions (such as two species that may respond to the
464 presence of a common predator, but do not directly interact) that may be useful in further
465 understanding potential fish species associations (*e.g.*, Clark et al. 2018; Popovic et al. 2019).

466 Although inferring species interactions from species associations is fraught with many
467 challenges, the combination of estimated species associations with empirical estimates of species
468 interaction strengths has the potential to provide novel insight into community associations

469 (Thurman et al. 2019). In addition, as Barner et al. (2018) suggest, although estimated species
470 associations may not capture true interactions, species associations may be useful in predicting
471 higher-level community properties such as the total number of interacting species, number of
472 interactions within/among trophic levels, and the proportion of positive and negative
473 interactions. Lastly, unlike residual correlations, the odds ratio metric used here (Gelfand and
474 Shirota, in review) generates interpretable measures of species co-occurrences that can be used to
475 better understand dependencies among species and how they vary spatially. Overall, JSDMs,
476 their derived metrics (*i.e.*, odds ratios), and complementary models such as Markov network
477 models provide the opportunity to gain a better understanding of the potential importance of
478 biotic interactions, how they vary across space, and their roles in structuring fish communities.

479 *Summary*

480 Understanding the effects of natural environmental and habitat conditions, anthropogenic
481 stressors, and management actions on the distribution of inland freshwater fishes is critically
482 important, especially in the context of global change (Araújo and Luoto 2007). JSDMs represent
483 a useful modeling framework that accounts for the multivariate nature of fish communities and
484 can accommodate a variety of data types (*e.g.*, presence/absence, abundance, ordinal counts) and
485 hierarchical, temporally/spatially autocorrelated data structures that are common in fish
486 community data (*i.e.*, through the use of random effects). These models can also be modified to
487 accommodate such issues as imperfect detection of fish within a community (Tobler et al. 2019).
488 Fortunately, a large number of JSDM approaches have been developed, many of which are easily
489 accessible and can be freely implemented in software such as the programming environment R
490 (see Norberg et al. 2019 and Wilkinson et al. 2019 for review of modeling approaches and
491 comparisons). The application of JSDMs in fisheries research represents an opportunity to

492 improve our understanding of the abiotic and biotic factors influencing the distribution of fish
493 communities, for predicting multispecies responses to changing environmental conditions, and
494 ultimately for improving management of multiple interacting fish species in the face of
495 environmental change.

496 **Acknowledgements**

497 Thanks to the many Minnesota DNR and PCA employees who have collected fish and
498 environmental data over the past several decades that were used in this analysis. Any use of
499 trade, firm, or product names is for descriptive purposes only and does not imply endorsement by
500 the U.S. Government.

Draft

502 **References**

- 503 Araújo, M.B., and Luoto, M. 2007. The importance of biotic interactions for modelling species
504 distributions under climate change. *Global Ecol. Biogeogr.* **16**(6): 743-753.
505 doi:10.1111/j.1466-8238.2007.00359.x.
- 506 Barner, A.K., Coblenz, K.E., Hacker, S.D, and Menge, B.A. 2018. Fundamental contradictions
507 among observational and experimental estimates of non-trophic species interactions.
508 *Ecology* **99**(3): 557-566. doi:10.1002/ecy.2133.
- 509 Bond, N., Thomson, J., Reich, P., and Stein, J. 2011. Using species distribution models to infer
510 potential climate change-induced range shifts of freshwater fish in south-eastern
511 Australia. *Mar. Freshwater Res.* **62**(9): 1043-1061. doi:10.1071/MF10286.
- 512 Bjørnstad, O.N. and Falck, W. 2001. Nonparametric spatial covariance functions: Estimating and
513 testing. *Env. Ecol. Stat.* 8:53-70.
- 514 Bjørnstad, O.N. 2020. ncf: Spatial covariance functions. R package version 1.2-9.
515 <https://CRAN.R-project.org/package=ncf>
- 516 Chezik, K.A., Lester, N.P., and Venturelli, P.A. 2014. Fish growth and degree-days I: Selecting a
517 base temperature for a within-population study. *Can. J. Fish. Aquat. Sci.* **71**(1): 47-55.
518 doi:10.1139/cjfas-2013-0295.
- 519 Chu, C., Mandrak, N.E., and Minns, C.K. 2005. Potential impacts of climate change on the
520 distributions of several common and rare freshwater fishes in Canada. *Divers. Distrib.*
521 **11**(4): 299-310. doi:10.1111/j.1366-9516.2005.00153.x.
- 522 Clark, J.S., Bell, D.M., Kwit, M.C., and Zhu, K. 2014a. Competition-interaction landscapes for
523 the joint response of forests to climate change. *Glob. Change Biol.* **20**(6): 1979-1991.
524 doi:10.1111/gcb.12425.

- 525 Clark, J.S., Nemergut, D., Seyednasrollah, B., Turner, P.J., and Zhang, S. 2017. Generalized
526 joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious
527 data. *Ecol. Monogr.* **87**(1): 34-56. doi:10.1002/ecm.1241.
- 528 Clark, N.J., Wells, K., and Lindberg, O. 2018. Unravelling changing interspecific interactions
529 across environmental gradients using Markov random fields. *Ecology* **99**(6): 1277-1283.
530 doi:10.1002/ecy.2221.
- 531 Clark, J.S., Gelfand, A.E., Woodall, C.W., Zhu, K. 2014b. More than the sum of the parts: forest
532 climate response from joint species distribution models. *Ecol. Appl.* **24**(5): 990-999.
533 doi:10.1890/13-1015.1.
- 534 DeWeber, J.T., and Wagner, T. 2018. Probabilistic measures of climate change vulnerability,
535 adaptation action benefits, and related uncertainty from maximum temperature metric
536 selection. *Glob. Change Biol.* **24**(6): 2735-2748. doi:10.1111/gcb.14101.
- 537 Dormann, C.F., Bobrowski, M., Dehling, D.M., Harris, D.J., Hartig, F., Lischke, H., Moretti,
538 M.D., Pagel, J., Pinkert, S., Schleuning, M., Schmidt, S.I., Sheppard, C.S., Steinbauer,
539 M.J., Zeuss, D., and Kraan, C. 2018. Biotic interactions in species distribution modeling:
540 10 questions to guide interpretation and avoid false conclusions. *Global Ecol. Biogeogr*
541 **27**(9): 1004-1016. doi:10.1111/geb.12759.
- 542 Gelfand, A.E., and Shirota, S. In review. Clarifying species dependence under joint species
543 distribution modeling. doi: 10.1101/744359.
- 544 Gibson-Reinemer, D. K., Stewart, D. R., Fritts, M. W., DeBoer, J. A., and Casper, A. F. 2016.
545 Estimating the effects of environmental variables and gear type on the detection and
546 occupancy of large-river fishes in a standardized sampling program using multiseason
547 bayesian mixture models. *N. Am. J. Fish. Manag.* 36:1445-1456.

- 548 Godefroid, M., Boldt, J.L., Thorson, J.T., Forrest, R., Gauthier, S., Flostrand, L., Perry, R.I.,
549 Ross, A.R.S., Galbraith, M. 2019. Spatio-temporal models provide new insights on. The
550 biotic and abiotic drivers shaping Pacific Herring (*Clupea pallasii*) distribution. Prog.
551 Oceanogr. 178:102198.
- 552 Gwinn, D. C., Beesley, L. S., Close, P., Gawne, B., & Davies, P. M. 2016. Imperfect detection
553 and the determination of environmental flows for fish: challenges, implications and
554 solutions. Freshw. Biol. 61:172-180.
- 555 Harris, D.J. 2015. Generating realistic assemblages with a joint species distribution model.
556 Methods Ecol. Evol. 6(4): 465-473. doi:10.1111/2041-210X.12332.
- 557 Hill, N.A., Foster, S.D., Duhamel, G., Welsford, D., Koubbi, P., and Johnson, C.R. 2017. Model-
558 based mapping of assemblages for ecology and conservation management: A case study
559 of demersal fish on the Kerguelen Plateau. Divers. Distrib. 23:1216-1230.
- 560 Hocking, D. J., Thorson, J. T., O'Neil, K., and Letcher, B. H. 2018. A geostatistical state-space
561 model of animal densities for stream networks. Ecol. Appl. 28:1782-1796.
- 562 Inoue, K., Stoeckl, K., and Geist, J. 2017. Joint species models reveal the effects of environment
563 on community assemblage of freshwater mussels and fishes in European rivers. Divers.
564 Distrib. 23(3): 284-296. doi:10.1111/ddi.12520.
- 565 Jacobson, P.C., Cross, T.K., Dustin, D.L., and Duval, M. 2016. A fish habitat conservation
566 framework for Minnesota Lakes. Fisheries 41(6): 302-317.
567 doi:10.1080/03632415.2016.1172482.
- 568 Kéry, M. 2011. Towards the modelling of true species distributions. J. Biogeogr. 38:617-618.

- 569 Latimer, A.M., Banerjee, S., Sang Jr, H., Mosher, E.S., and Silander Jr, J.A. 2009. Hierarchical
570 models facilitate spatial analysis of large data sets: a case study on invasive plant species
571 in the northeastern United States. *Ecol. Lett.* 12:144-154.
- 572 MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., A. J. Royle and Langtimm, C. A.
573 2002. Estimating site occupancy rates when detection probabilities are less than one.
574 *Ecology* 83: 2248-2255.
- 575 Manel S., Williams H.C., and Ormerod S.J. 2001 Evaluating presence–absence models in
576 ecology: the need to account for prevalence. *J. Appl. Ecol.* **38**:921–931.
- 577 MNDNR. 2017. Manual of instructions for lake survey. Division of Fish and Wildlife, Section of
578 Fisheries, Special Publication 180. St. Paul, MN.
- 579 Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J., Anttila, J., et al. 2019. A
580 comprehensive evaluation of predictive performance of 33 species distribution models at
581 species and community levels. *Ecol. Monogr.* **89**(3): e01370. doi:10.1002/ecm.1370.
- 582 Ovaskainen, O., Tikhonov, G., Norberg, A., Blanchet, F.G., Duan, L., Dunson, D., Roslin, T.,
583 and Abrego, N. 2017. How to make more out of community data? A conceptual
584 framework and its implementation as models and software. *Ecol. Lett.* **20**(5): 561-576.
585 doi:10.1111/ele.12757.
- 586 Pendleton, R.M., Pritt, J.J., Peoples, B.K., and Frimpong, E.A. 2012. The strength of Nocomis
587 nest association contributes to patterns of rarity and commonness among New River,
588 Virginia cyprinids. *Am. Midl. Nat.* **168**(1): 202-217. doi: 10.1674/0003-0031-168.1.202.
- 589 Peoples, B.K., Blanc, L.A., and Frimpong, E.A. 2015. Lotic cyprinid communities can be
590 structures as nest webs and predicted by the stress-gradient hypothesis. *J. Anim. Ecol.*
591 **84**(6): 1666-1677. doi:10.1111/1365-2656.12428.

- 592 Poff, N.L. 1997. Landscape filters and species traits: towards mechanistic understanding and
593 prediction in stream ecology. *J. N. Am. Benthol. Soc.* **16**(2): 391-409.
594 doi:10.2307/1468026.
- 595 Poole, H.H., and Atkins, W.R.G. 1929. Photo-electric measurements of submarine illumination
596 throughout the year. *J. Mar. Biol. Assoc. U.K.* **16**(1): 297-324.
597 doi:10.1017/S0025315400029828.
- 598 Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A.,
599 and McCarthy, M.A. 2014. Understanding co-occurrence by modelling species
600 simultaneously with a Joint Species Distribution Model (JSDM). *Methods Ecol. Evol.*
601 **5**(5): 397-406. doi:10.1111/2041-210X.12180.
- 602 Popovic, G.C., Warton, D.I., Thomson, F.J., Hui, F.K.C. 2019. Untangling direct species
603 associations from indirect mediator species effects with graphical models. *Methods Ecol.*
604 *Evol.* **10**(9): 1571-1583. doi: 10.1111/2041-210X.13247.
- 605 Prabhakaran, S. 2016. InformationValue: Performance Analysis and Companion Functions for
606 Binary Classification Models. R package version 1.2.3. [https://CRAN.R-](https://CRAN.R-project.org/package=InformationValue)
607 [project.org/package=InformationValue](https://CRAN.R-project.org/package=InformationValue)
- 608 R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for
609 Statistical Computing, Vienna, Austria
- 610 Radinger, J., Alcaraz-Hernández, J.D., and García-Berthou, E. 2019. Environmental filtering
611 governs the spatial distribution of alien fishes in a large, human-impacted Mediterranean
612 river. *Divers. Distrib.* **25**(5): 701-714. doi:10.1111/ddi.12895.

- 613 Roy, A.H., Freeman, B.J., Freeman, M.C. 2007. Riparian influences on stream fish assemblage
614 structure in urbanizing streams. *Landsc. Ecol.* **22**(3):385-402. doi:10.1007/s10980-006-
615 9034-x.
- 616 Stewart, D.R., Walters, A.W., Rahel, F.J. 2016. Landscape-scale determinants of native and non-
617 native Great Plains fish distributions. *Divers. Distrib.* **22**(2): 225-238. doi:
618 10.1111/ddi.12383.
- 619 Stjernman, M., Sahlin, U., Olsson, O., and Smith, H.G. 2019. Estimating effects of arable land
620 use intensity on farmland birds using joint species modeling. *Ecol. Appl.* **29**(4):e01875.
621 doi: 10.1002/eap.1875.
- 622 Sutherland, A.B., Meyer, J.L., and Gardiner, E.P. 2002. Effects of land cover on sediment regime
623 and fish assemblage structure in four southern Appalachian streams. *Freshw. Biol.* **47**(9):
624 1791–1805. doi:10.1046/j.1365-2427.2002.00927.x.
- 625 Taylor, A.T., Papes, M., and Long, J.M. 2018. Incorporating fragmentation and non-native
626 species into distribution models to inform fluvial fish conservation. *Conserv. Biol.* **32**(1):
627 171-182. doi: 10.1111/cobi.13024.
- 628 Thorson, J.T. Scheuerell, M.D., Shelton, A.O., See, K.E., Skaug, H.J. and Kristensen, K. 2015.
629 Spatial factor analysis: a new tool for estimating joint species distributions and
630 correlations in species range. *Meth. Ecol. Evol.* **6**:627-637.
- 631 Thorson, J.T., Ianelli, J.N., Larsen, E.A., Ries, L., Scheuerell, M.D., Szuwalski, C., and Zipken,
632 E.F. 2016. Joint dynamic species distribution models: a tool for community ordination
633 and spatio-temporal monitoring. *Global Ecol. Biogeogr.* **25**:1144-1158.
- 634 Thorson, J.T., Munch, S.B., and Swain, D.P. 2017. Estimating partial regulation in
635 spatiotemporal models of community dynamics. *Ecology* **98**:1277-1289.

- 636 Thorson, J.T. 2019. Guidance for decisions using the Vector Autoregressive Spatio-temporal
637 (VAST) package in stock, ecosystem, habitat, and climate assessments. *Fish. Res.*
638 210:143-161.
- 639 Thurman, L.L., Barner, A.K., Garcia, T.S., and Chestnut, T. 2019. Testing the link between
640 species interactions and species co-occurrence in a trophic network. *Ecography* 42:1-13.
- 641 Tikhonov, G., Abrego, N., Dunson, D., and Ovaskainen, O. 2017. Using joint species
642 distribution models for evaluating how species-to-species associations depend on the
643 environmental context. *Methods Ecol. Evol.* 8:443-452.
- 644 Tobler, M W., M. Kéry, F.K.C. Hui, G. Guillera-Aroita, P. Knaus, and T. Sattler. 2019. Joint
645 species distribution models with species correlations and imperfect detection. *Ecology.*
- 646 Trebitz, A., Carpenter, S., Cunningham, P., Johnson, B., Lillie, R., Marshall, D., Martin, T.,
647 Narf, R., Pellet, T., Stewart, S., Storlie, C., and Unmuth, J. 1997. A model of bluegill-
648 largemouth bass interactions in relation to aquatic vegetation and its management. *Ecol.*
649 *Model.* **94**(2-3):139-156. doi:10.1016/S0304-3800(96)00009-9.
- 650 Van Zuiden, T., and Sharma., S. 2016. Examining the effects of climate change and species
651 invasions on Ontario walleye populations: can walleye beat the heat? *Divers. Distrib.*
652 **22**(10): 1069-1079. doi:10.1111/ddi.12468.
- 653 Ver Hoef, J. M., Peterson, E., and Theobald, D. 2006. Spatial statistical models that use flow and
654 stream distance. *Environ. Ecol. Stat.* 13:449-464.
- 655 Wagner, T., Deweber, J.T., Detar, J., and Sweka, J.A. 2013. Landscape-scale evaluation of
656 asymmetric interactions between brown trout and brook trout using two-species
657 occupancy models. *Trans. Am. Fish. Soc.* **142**(2): 353–361.
658 doi:10.1080/00028487.2012.734892

- 659 Wang, L., Lyons, J., Kanehl, P., and Bannerman, R. 2001. Impacts of urbanization on stream
660 habitat and fish across multiple spatial scales. *Environ. Manage.* **28**(2): 255-266.
661 doi:10.1007/s0026702409.
- 662 White, S.L., and Orth, D.J. 2013. Distribution and habitat correlates of Clinch Dace (*Chrosomus*
663 sp. cf. *saylori*) in the upper Clinch River watershed. *Am. Midl. Naturalist.* **171**(2): 311-
664 321. doi:10.1674/0003-0031-171.2.311.
- 665 White, S.L., and Orth, D.J. 2014. Reproductive biology of Clinch dace, *Chrosomus* sp. cf.
666 *saylori*. *Southeast. Nat.* **13**(4): 735-744. doi:10.1656/058.013.0404.
- 667 White, S., Faulk, E., Tzilkowski, C., Weber, A.S., Marshall, M., and Wagner, T. 2020.
668 Predicting fish species richness and habitat relationships using Bayesian hierarchical
669 multispecies occupancy models. *Can. J. Fish. Aquat. Sci.* **77**: 602-610.
- 670 Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R., and McCarthy, M.A. 2019. A
671 comparison of joint species distribution models for presence-absence data. *Methods Ecol.*
672 *Evol.* **10**(2):198-211. doi: 10.1111/2041-210X.13106.
- 673 Winslow, L.A., Hansen, G.J.A., Read, J.S., and Notaro, M. 2017. A large-scale database of
674 modeled contemporary and future projected water temperature data for 10,774
675 Midwestern U.S. Lakes. *Scientific Data* 4: 170053.
- 676 Xenopoulos, M.A., Lodge, D.M., Alcamo, J., Märker, M., Schulze, K., and Van Vuuren, D.P.
677 2005. Scenarios of freshwater fish extinctions from climate change and water withdrawal.
678 *Glob. Change Biol.* **11**(10): 1557–1564. doi:10.1111/j.1365-2486.2005.001008.x.
- 679 Zurell, D., Pollock, L.J., and Thuiller, W. 2018. Do joint species distribution models reliably
680 detect interspecific interactions from co-occurrence data in homogenous environments?.
681 *Ecography* **41**(11):1812-1819. doi:10.1111/ecog.03315.

683 **Figure captions**

684 **Figure 1.** Effects of environmental covariates on the occurrence of Appalachian stream fishes.

685 Circles are posterior means and bars are 95% credible intervals (CI). Blue symbols represent
686 those effects with 95% CIs that did not overlap zero.

687 **Figure 2.** Residual correlations and shared environmental correlations among species pairs for
688 Appalachian stream fishes. Only those correlations where the 95% CI of either residual or shared
689 environmental correlations did not overlap zero are shown. Colors signify the direction (positive
690 or negative) and type (environmental or residual) of significant correlation. Shapes signify the
691 type of significant correlations between species.

692 **Figure 3.** Posterior mean conditional (blue circles) and marginal (red circles) AUC values for
693 predicting occurrence of Appalachian stream fishes. Horizontal bars are 95% credible intervals.
694 Shading indicates AUC thresholds separating models of low accuracy ($AUC < 0.7$, white),
695 moderate accuracy ($0.7 < AUC < 0.9$; light grey), and high accuracy ($AUC > 0.9$, dark grey).

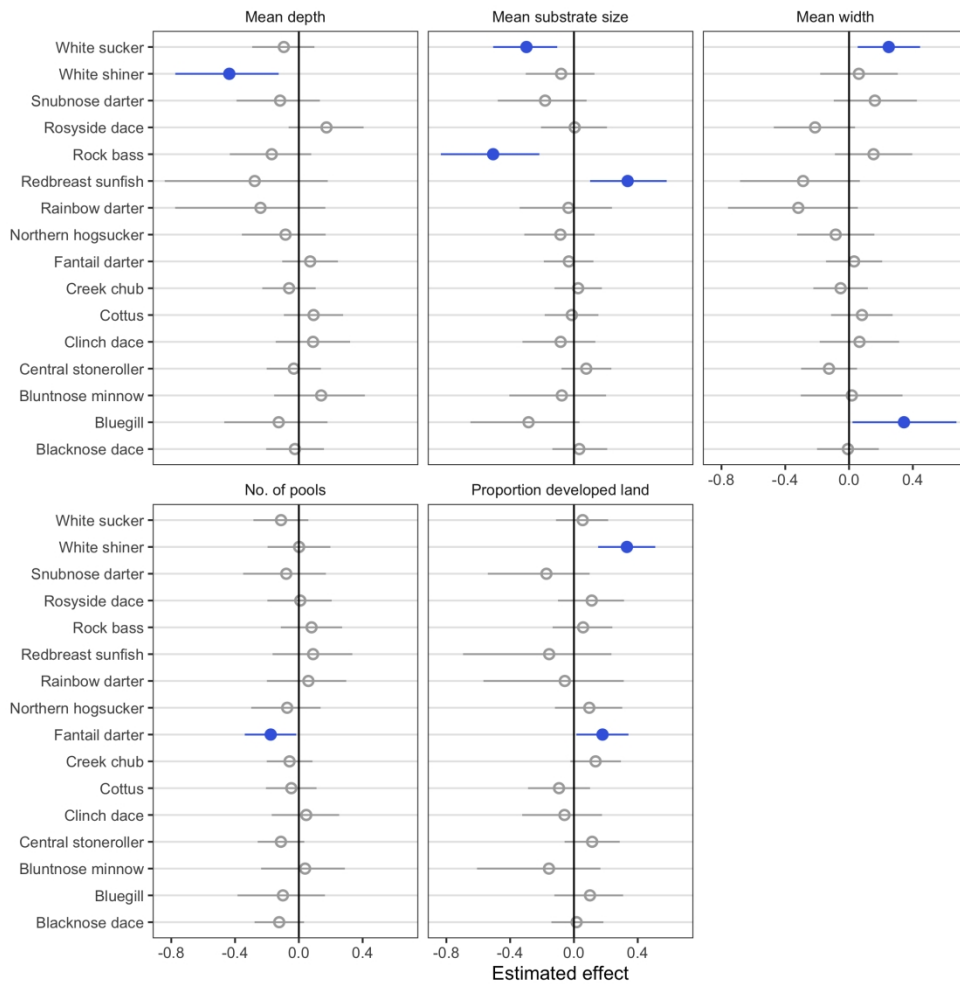
696 **Figure 4.** Effects of environmental covariates on the occurrence of north temperate lake fishes.
697 Circles are posterior means and bars are 95% credible intervals (CI). Blue symbols represent
698 those effects with 95% CIs that did not overlap zero.

699 **Figure 5.** Residual correlations and shared environmental correlations among species pairs for a
700 subset of north temperate lake fishes. Only those correlations where the 95% CI of either residual
701 or shared environmental correlations did not overlap zero are shown. Colors signify the direction
702 (positive or negative) and type (environmental or residual) of significant correlation. Shapes
703 signify the type of significant correlations between species. A subset of species is shown here for
704 clarity, see Supplementary Figure S1 for all species.

705 **Figure 6.** Log odds ratios for three species pairs (bluegill and largemouth bass, burbot and
706 golden shiner, and smallmouth bass and walleye) for each lake. Log odds ratios greater than zero
707 indicates positive dependence; values less than zero indicate negative dependence in species
708 occurrence. Vertical line at zero represents independence between species.

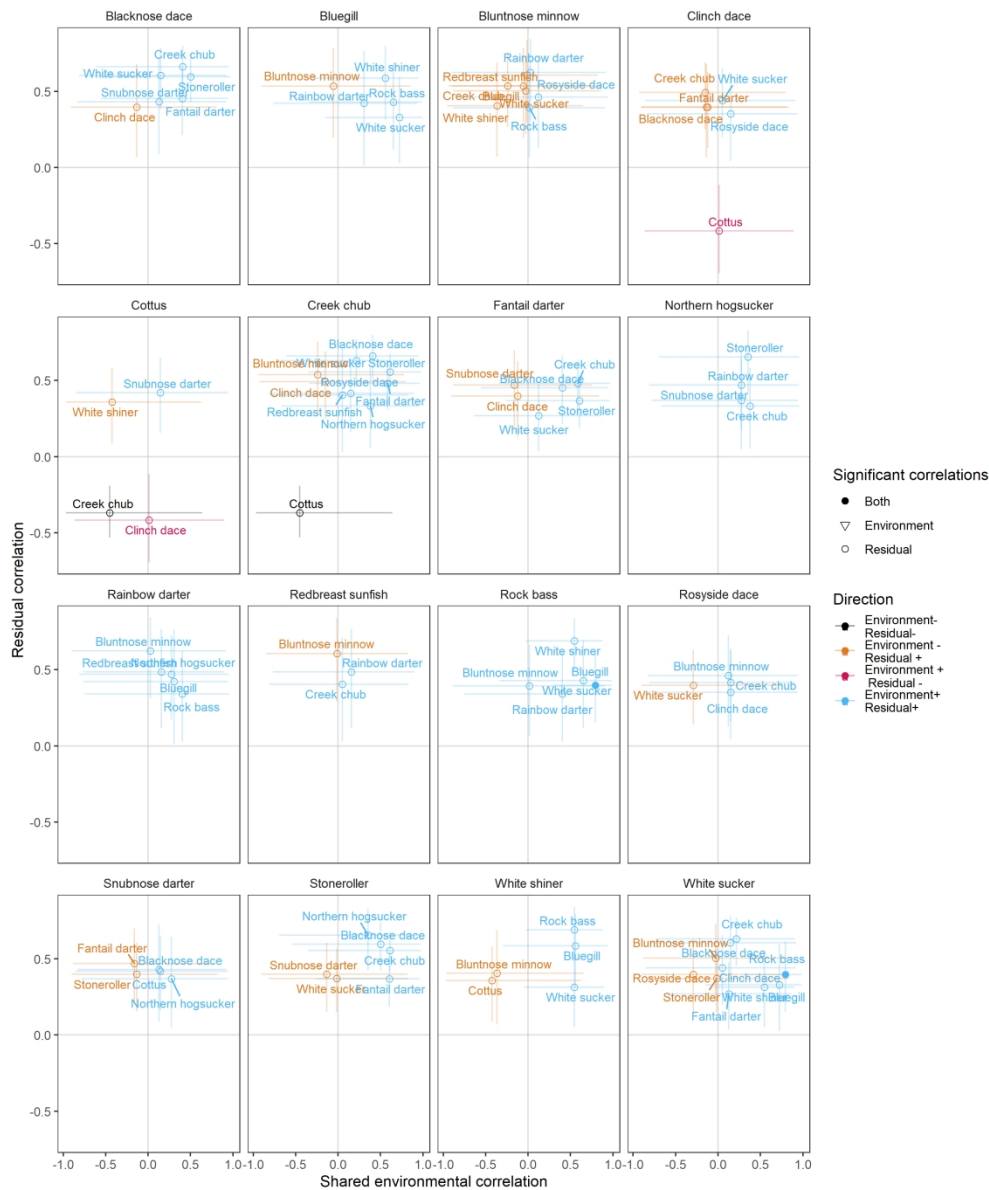
709 **Figure 7.** Posterior mean conditional (blue circles) and marginal (red circles) AUC values for
710 predicting occurrence of north temperate lake fishes. Horizontal bars are 95% credible intervals.
711 Shading indicates AUC thresholds separating models of low accuracy ($AUC < 0.7$, white),
712 moderate accuracy ($0.7 < AUC < 0.9$; light grey), and high accuracy ($AUC > 0.9$, dark grey).

713 **Figure 8.** Specificity and sensitivity for predicting the occurrence of north temperate lake fishes
714 according to prevalence in the dataset. Circles are posterior means and vertical bars are 95%
715 credible intervals, for marginal (blue) and conditional (red) predictions.

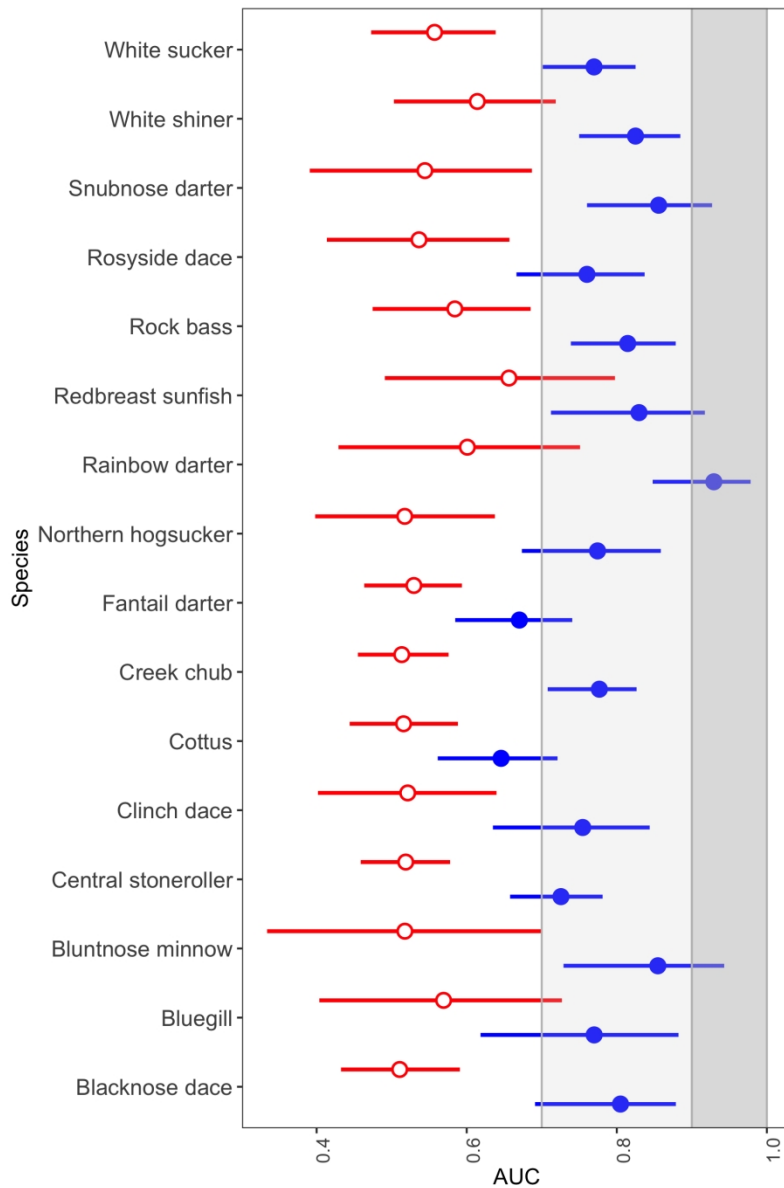


Effects of environmental covariates on the occurrence of Appalachian stream fishes. Circles are posterior means and bars are 95% credible intervals (CI). Blue symbols represent those effects with 95% CIs that did not overlap zero.

846x846mm (72 x 72 DPI)

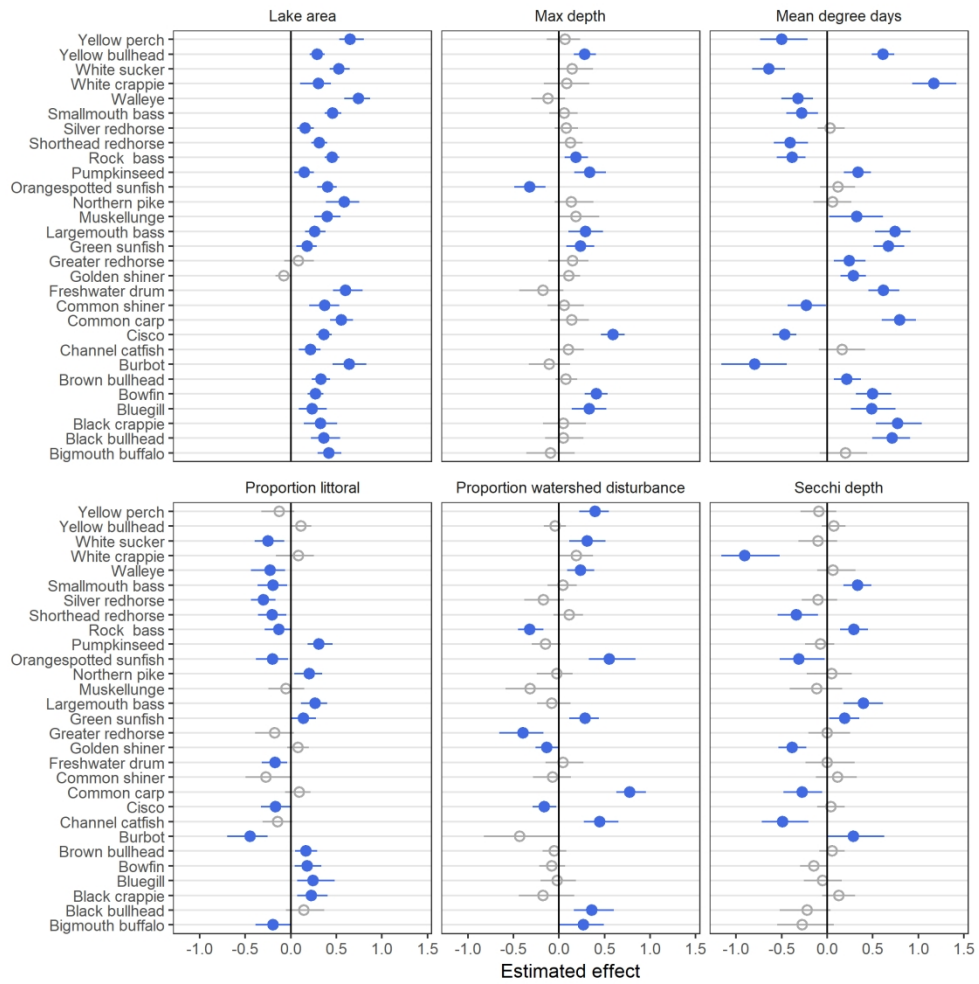


Residual correlations and shared environmental correlations among species pairs for Appalachian stream fishes. Only those correlations where the 95% CI of either residual or shared environmental correlations did not overlap zero are shown. Colors signify the direction (positive or negative) and type (environmental or residual) of significant correlation. Shapes signify the type of significant correlations between species.

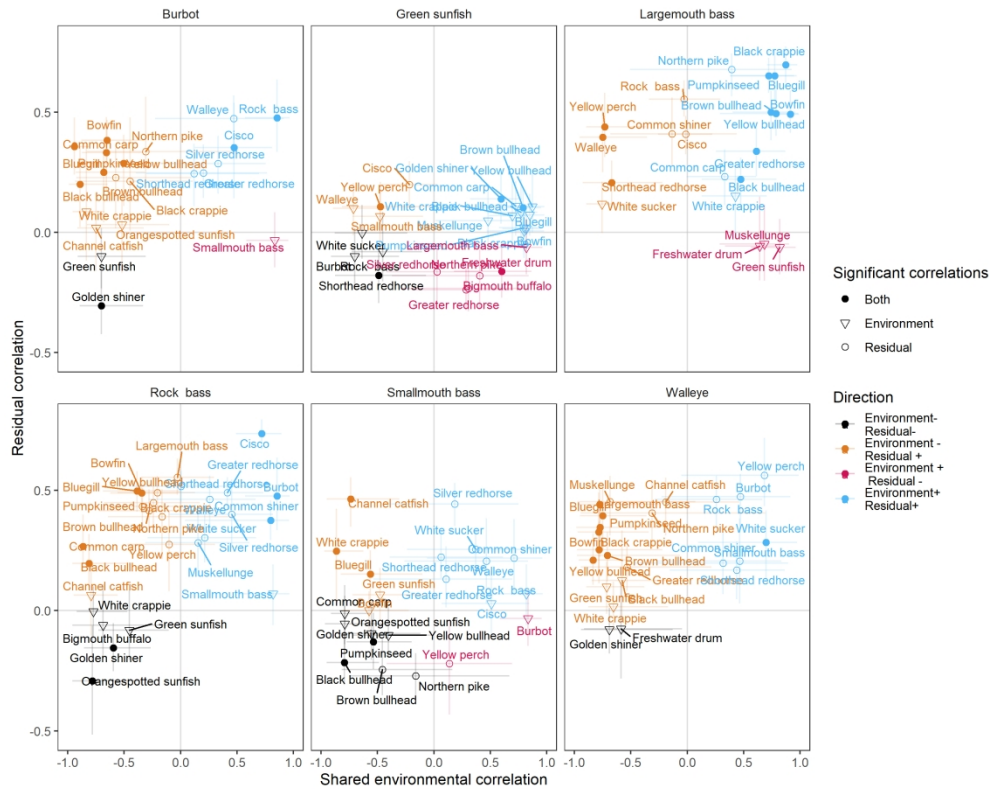


Posterior mean conditional (blue circles) and marginal (red circles) AUC values for predicting occurrence of Appalachian stream fishes. Horizontal bars are 95% credible intervals. Shading indicates AUC thresholds separating models of low accuracy ($AUC < 0.7$, white), moderate accuracy ($0.7 < AUC < 0.9$; light grey), and high accuracy ($AUC > 0.9$, dark grey).

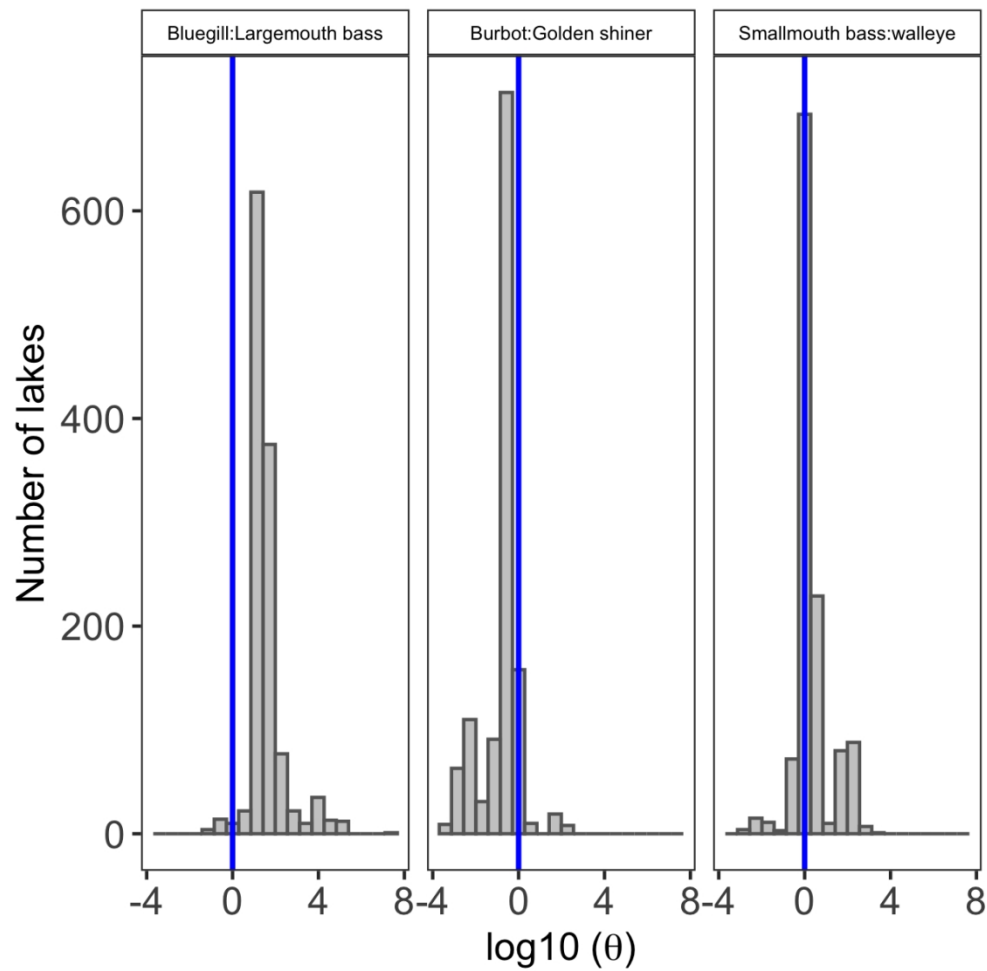
635x952mm (72 x 72 DPI)



Effects of environmental covariates on the occurrence of north temperate lake fishes. Circles are posterior means and bars are 95% credible intervals (CI). Blue symbols represent those effects with 95% CIs that did not overlap zero.

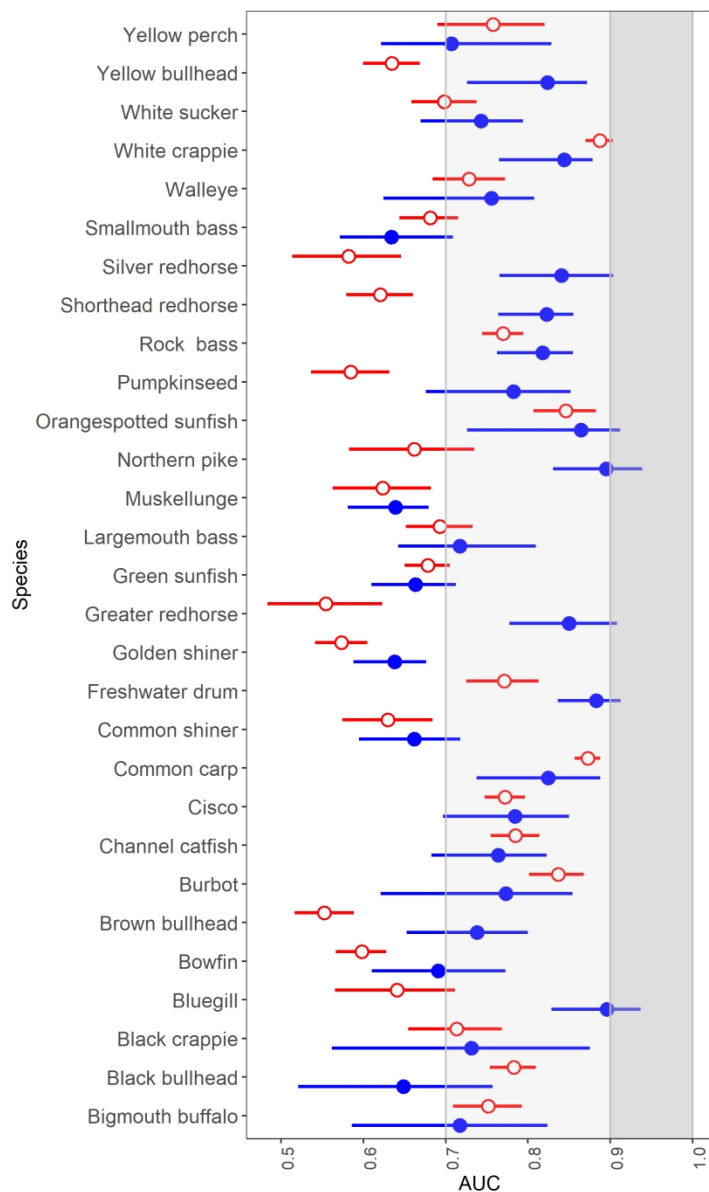


Residual correlations and shared environmental correlations among species pairs for a subset of north temperate lake fishes. Only those correlations where the 95% CI of either residual or shared environmental correlations did not overlap zero are shown. Colors signify the direction (positive or negative) and type (environmental or residual) of significant correlation. Shapes signify the type of significant correlations between species. A subset of species is shown here for clarity, see Supplementary Figure S1 for all species.

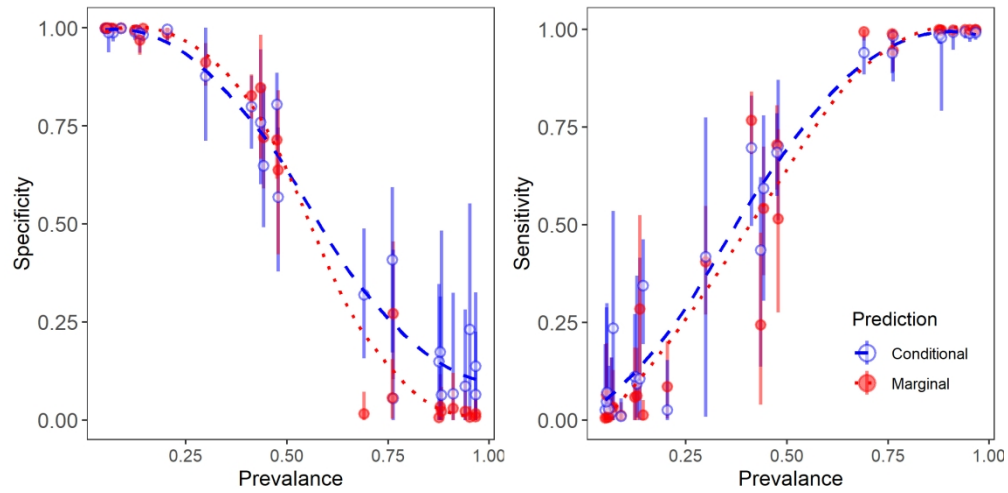


Log odds ratios for three species pairs (bluegill and largemouth bass, burbot and golden shiner, and smallmouth bass and walleye) for each lake. Log odds ratios greater than zero indicates positive dependence; values less than zero indicate negative dependence in species occurrence. Vertical line at zero represents independence between species.

458x458mm (72 x 72 DPI)



Posterior mean conditional (blue circles) and marginal (red circles) AUC values for predicting occurrence of north temperate lake fishes. Horizontal bars are 95% credible intervals. Shading indicates AUC thresholds separating models of low accuracy (AUC<0.7, white), moderate accuracy (0.7<AUC<0.9; light grey), and high accuracy (AUC>0.9, dark grey).



Specificity and sensitivity for predicting the occurrence of north temperate lake fishes according to prevalence in the dataset. Circles are posterior means and vertical bars are 95% credible intervals, for marginal (blue) and conditional (red) predictions.