

WILEY

ECOLOGICAL APPLICATIONS

**A model-based assessment of anthropogenic disturbance on
lotic macroinvertebrate assemblages**

Journal:	<i>Ecological Applications</i>
Manuscript ID	Draft
Wiley - Manuscript type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Kopp, Darin; Oak Ridge Institute for Science and Education; United States Environmental Protection Agency, Office of Research and Development Stoddard, John; U.S. Environmental Protection Agency Kaufmann, Philip; Environmental Protection Agency, Office of Research and Development; Oregon State University, Fisheries & Wildlife Herlihy, Alan; Oregon State University, Department of Fisheries, Wildlife and Conservation Science Hill, Ryan; United States Environmental Protection Agency, Office of Research and Development Brehob, Meredith; Oak Ridge Institute for Science and Education Sabo, Robert; US EPA Office of Research and Development
Substantive Area:	Community Ecology < Substantive Area, Disturbance < Community Ecology < Substantive Area, Statistics and Modeling < Theory < Substantive Area, Conservation < Landscape < Substantive Area
Organism:	Invertebrates < Animals, Insects < Arthropods < Invertebrates < Animals
Habitat:	Freshwater < Aquatic Habitat < Habitat
Geographic Area:	United States < North America < Geographic Area
Key words/phrases:	Reference Condition, Biological Assessment, Biodiversity, Macroinvertebrates, Joint Species Distribution Modeling
Abstract:	Traditionally, the effects of anthropogenic disturbance on biological assemblages are elucidated by comparing an assemblage observed at a site to one that represents a minimally disturbed state. Unfortunately, defining a minimally disturbed state is extremely challenging because of the extent of human disturbance. We use a national scale dataset and a two-stage model-based approach to assess how benthic macroinvertebrate assemblages at 1,748 sites would change if common anthropogenic disturbances were removed from in-stream physiochemical variables. First, we used random forest models and current landscape data to predict physiochemical conditions and then infer abiotic condition in the absence of disturbance. Second, we combined these estimates with joint species distribution models to predict the assemblage that would

occur in these undisturbed conditions. Random forest models explained 48 – 75% of the variation in total nitrogen, phosphorous, sulfate, chloride, and substrate diameter. Generally, nutrient and salinity concentrations were higher, and substrates were finer than predicted to be without disturbances. Using this physiochemical data, joint species distribution models accurately explained genus richness ($R^2 = 0.73 - 0.85$) and composition (Jaccard similarity index = 0.48 – 0.55). Depending on the ecoregion, we found that genus richness could change at 26 – 61% of sites if disturbance was removed. Different responses were observed for insect and non-insect taxa. For example, under anthropogenic disturbance, occurrence probabilities for Ephemeroptera, Plecoptera and Trichoptera tended to decrease at 5 – 26% of sites while occurrence probabilities for Mollusca and other non-insect, non-arthropod taxa increased at 5 – 33% and 11 – 24% of sites, respectively. Importantly, our framework provides an avenue for evaluating the effects of anthropogenic disturbance on macroinvertebrate assemblages without identifying reference sites.

SCHOLARONE™
Manuscripts

1 A model-based assessment of anthropogenic disturbance on lotic macroinvertebrate assemblages

2

3 Darin A. Kopp^{1*}, John L. Stoddard¹, Philip R. Kaufmann^{1,2}, Alan T. Herlihy², Ryan A. Hill¹,

4 Meredith M. Brehob³, Robert D Sabo⁴

5

6 ¹US EPA, Office of Research and Development, Corvallis, Oregon, USA

7 ²Department of Fisheries, Wildlife, & Conservation Sciences, Oregon State University,

8 Corvallis, Oregon, USA

9 ³Oak Ridge Institute for Science and Education Participant c/o US EPA, ORD, Washington,
10 D.C., USA

11 ⁴US EPA, Office of Research and Development, Center for Public Health and Environmental
12 Assessment, Washington, DC, USA

13

14 *Correspondence: kopp.darin@epa.gov

15

16 Keywords: Joint Species Distribution Modeling, Richness, National Rivers and Streams
17 Assessment, Benthic Macroinvertebrate, Reference Sites, National Aquatic Resource Surveys.

18

19 Open research statement: Benthic macroinvertebrate occurrence data and physiochemical
20 variables were retrieved from the 2018-2019 National Rivers and Stream Assessment
21 (<https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys>)
22 Benthic macroinvertebrate data can be downloaded as .csv file from [NRSA 1819 Benthic](#)
23 [Macroinvertebrate Count - Data \(CSV\) \(csv\)](#). Physiochemical data can be downloaded as .csv
24 files from [NRSA 1819 Water Chemistry_CHLA - Data \(CSV\) \(csv\)](#) and [NRSA 1819 Physical](#)
25 [Habitat Larger Set of Metrics - Data \(CSV\) \(csv\)](#) Landscape variables were retrieved from
26 <https://www.epa.gov/national-aquatic-resource-surveys/streamcat-metrics-and-definitions> and
27 PRISM Climate Group <https://prism.oregonstate.edu/>

28

29 Abstract:

30 Traditionally, the effects of anthropogenic disturbance on biological assemblages are
31 elucidated by comparing an assemblage observed at a site to one that represents a minimally
32 disturbed state. Unfortunately, defining a minimally disturbed state is extremely challenging
33 because of the extent of human disturbance. We use a national scale dataset and a two-stage
34 model-based approach to assess how benthic macroinvertebrate assemblages at 1,748 sites would
35 change if common anthropogenic disturbances were removed from in-stream physiochemical
36 variables. First, we used random forest models and current landscape data to predict
37 physiochemical conditions and then infer abiotic condition in the absence of disturbance.
38 Second, we combined these estimates with joint species distribution models to predict the
39 assemblage that would occur in these undisturbed conditions. Random forest models explained
40 48 – 75% of the variation in total nitrogen, phosphorous, sulfate, chloride, and substrate
41 diameter. Generally, nutrient and salinity concentrations were higher, and substrates were finer
42 than predicted to be without disturbances. Using this physiochemical data, joint species
43 distribution models accurately explained genus richness ($R^2 = 0.73 – 0.85$) and composition
44 (Jaccard similarity index = 0.48 – 0.55). Depending on the ecoregion, we found that genus
45 richness could change at 26 – 61% of sites if disturbance was removed. Different responses were
46 observed for insect and non-insect taxa. For example, under anthropogenic disturbance,
47 occurrence probabilities for Ephemeroptera, Plecoptera and Trichoptera tended to decrease at 5 –
48 26% of sites while occurrence probabilities for Mollusca and other non-insect, non-arthropod
49 taxa increased at 5 – 33% and 11 – 24% of sites, respectively. Importantly, our framework
50 provides an avenue for evaluating the effects of anthropogenic disturbance on macroinvertebrate
51 assemblages without identifying reference sites.

52 Introduction:

53 Quantifying the extent and magnitude of anthropogenic disturbance on ecosystems
54 requires a benchmark that represents a desired, expected, or previous condition (Stoddard et al.
55 2006, Hawkins et al. 2010, McNellie et al. 2020). Ideally, benchmarks should represent
56 chemical, physical, and biological characteristics with minimal anthropogenic disturbance
57 (Reynoldson et al. 1997, Stoddard et al. 2006). However, for many ecosystems the minimally
58 disturbed condition is difficult to quantify or may no longer exist (Dudgeon et al. 2006,
59 Vörösmarty et al. 2010, McNellie et al. 2020). The concept of “least disturbed” describes
60 conditions that are derived from a collection of sites that represent the best available, given
61 today’s landscape (Stoddard et al. 2006). Unfortunately, because anthropogenic disturbance is
62 not evenly distributed across the landscape, sites in least disturbed condition can be spatially
63 biased and/or vary in quality (Herlihy et al. 2008, McNellie et al. 2020, Yuan et al. 2024). If least
64 disturbed sites are spatially biased, their ecological, chemical, and physical conditions may be
65 unattainable for some locations that need to be assessed because of differences in natural setting
66 or biogeographic history. Alternatively, if reference site quality is lower for some ecosystems
67 because of a relatively long legacy of human intensification, some sites will be judged against a
68 lower benchmark and erroneously appear to be in better condition relative to others. Although
69 least disturbed reference sites provides reasonable and defensible benchmarks for assessing
70 biological condition (Stoddard et al. 2006, Herlihy et al. 2008, Mitchell et al. 2025), they are
71 typically not undisturbed, nor evenly distributed among all ecosystem types (McNellie et al.
72 2020, Yuan et al. 2024). Novel approaches and concepts are urgently needed to overcome these
73 limitations.

74 Quantifying taxon-environment relationships for all taxa within a region with empirical
75 models may provide an alternative to relying on reference sites because they elucidate how
76 individual taxa and entire assemblages change along physiochemical gradients (Chessman and
77 Royal 2004, Kilgour and Stanfield 2005, Elias et al. 2016, Kopp et al. 2023). Because
78 anthropogenic disturbance typically manifests as altered physiochemical conditions (Tang et al.
79 2020, Fergus et al. 2023), quantifying potential environmental divers across space, could be
80 relevant for understanding how organisms may change with disturbance across time (Blüthgen et
81 al. 2022). For example, taxa that have a negative relationship along an environmental gradient
82 could be expected to become less prevalent if human disturbance contributes to more extreme
83 environmental conditions and more prevalent if reducing human disturbance contributes to less
84 extreme conditions (Kopp et al. 2023). From these relationships it is possible to predict which
85 taxa from the regional species pool could occur at a site if it were minimally disturbed and
86 compare them to the taxa that were actually observed (Chessman and Royal 2004, Elias et al.
87 2016). If the minimally disturbed assemblage is similar to the observed assemblage, then the site
88 may be considered unaffected by anthropogenic disturbances.

89 Advances in species distribution modeling enhance our ability to quantify taxon-
90 environment relationships (Franklin 2010, Guisan et al. 2017, Ovaskainen and Abrego 2020).
91 Species distribution modeling consists of a diverse set of methods that typically focus on
92 quantifying taxon-environment relationships for individual species (Elith and Leathwick 2009).
93 Joint species distribution modeling (JSDM) is a recent multivariate extension of single species
94 distribution modeling, capable of estimating taxon-environment relationships for all members of
95 an assemblage simultaneously. Because of their complex structure, JSDMs are often fitted using
96 Bayesian inference. As such, the resulting posterior distribution can be used for hypothesis

97 testing (Johnson et al. 2022). For example, given a posterior distribution of values representing a
98 biological assemblage (e.g. taxonomic richness) in the absence of human disturbance, it is
99 possible to evaluate whether an observed assemblage is consistent with this distribution. In
100 addition, these models use latent factors to account for unmeasured environmental variables and
101 potential biotic interactions (Warton et al. 2015, Ovaskainen et al. 2017, Ovaskainen and Abrego
102 2020). This feature is particularly important for biological assessments because it can strengthen
103 the inference gained from evaluating a specific suite of environmental variables that are typically
104 altered by human disturbance. Thus, the ability to model all taxa in an assemblage and
105 probabilistically evaluate whether an observation is consistent with a posterior distribution
106 makes JSDMs well suited for understanding how anthropogenic factors affect biological
107 assemblages.

108 Although taxon-environment relationships are necessary to circumvent the need for
109 reference sites, they require a concurrent expectation of undisturbed physiochemical conditions
110 to fully understand the effects of anthropogenic activities. Traditionally, undisturbed
111 physiochemical conditions are obtained, or modeled, from least-disturbed reference sites (Olson
112 and Hawkins 2012, Olson and Cormier 2019). However, empirical models have also linked
113 changes in the physiochemical environment to anthropogenic activities and been used to estimate
114 conditions if anthropogenic disturbance is reduced or eliminated without reference sites (Dodds
115 and Oakes 2004, Herlihy and Sifneos 2008, Soranno et al. 2011). Random forest models are a
116 machine learning algorithm that have superior predictive performance compared to other
117 modeling techniques (Prasad et al. 2006, Peters et al. 2007) and are often used to model water
118 quality parameters and physical habitat characteristics (Olson and Hawkins 2013, Olson and
119 Cormier 2019, Sabo et al. 2023, Yuan et al. 2024). Further these models are more robust to

120 nonlinearities, multicollinearity, and overfitting (Breiman 2001) and can evaluate the relative
121 importance of different predictor variables (Lin et al. 2021, Sabo et al. 2023). Because these
122 algorithms can accommodate complex relationships and a relatively high number of covariates,
123 they are well suited for evaluating the relative importance of natural and anthropogenic variables
124 on physiochemical conditions and, in turn, predicting the abiotic conditions if anthropogenic
125 disturbances were removed (Yuan et al. 2024). Indeed, efforts that combined estimates of
126 undisturbed physiochemical conditions with biological models have made progress towards
127 addressing the limitations of reference site-based approaches (Chessman and Royal 2004,
128 Kilgour and Stanfield 2005, Elias et al. 2016, Yuan et al. 2024).

129 In rivers and streams, assessments of biological integrity often rely on surveys of benthic
130 macroinvertebrate assemblages because they are diverse, relatively easy to sample, and respond
131 to changes in the physiochemical environment associated with anthropogenic disturbance
132 (Hughes and Peck 2008, Buss et al. 2014). To elucidate the effects of anthropogenic activities on
133 biotic integrity, benthic macroinvertebrate assemblages observed at a site are typically compared
134 to reference assemblages collected from least disturbed sites (Hawkins 2006, Stoddard et al.
135 2008). Here, we develop a model-based assessment as a complementary method to evaluate how
136 benthic macroinvertebrate assemblages at 1,748 stream sites may change if the effects of major
137 anthropogenic disturbances were removed from the physiochemical environment. First, we use
138 random forest models to relate in-stream physiochemical conditions to a suite of geoclimatic and
139 anthropogenic factors and use these models to infer abiotic conditions if anthropogenic
140 disturbance was eliminated. Second, we combined these estimates with JSMDs to predict how
141 biological assemblages could change if physiochemical conditions were not altered by
142 anthropogenic disturbance. Because JSMDs can leverage latent factors to account for

143 unmeasured environmental factors, including potential biotic interactions (Ovaskainen et al.
144 2016), we focus exclusively on environmental variables that are commonly associated with
145 anthropogenic disturbances. Specifically, we evaluate the effects of changing nutrient and
146 salinity concentrations, physical habitat, and climate on benthic macroinvertebrate assemblages
147 because these gradients are commonly used as abiotic screens used to identify least disturbed
148 reference sites (Herlihy et al. 2008, Paulsen et al. 2008, USEPA 2023, Mitchell et al. 2025) and
149 affect benthic macroinvertebrate distributions (Kopp et al. 2023).

150

151 Methods:

152 For our analysis, we focused on biological and environmental data collected at sites
153 distributed among 9 ecoregions (Figure 1) and surveyed during the 2018/2019 National Rivers
154 and Streams Assessment (NRSA). NRSA is a collaboration between the United States
155 Environmental Protection Agency (USEPA) and state, tribal, and federal partners to assess the
156 chemical, physical, and biological condition of streams and rivers of the United States
157 (<https://www.epa.gov/national-aquatic-resource-surveys/nrsa>). Every five years, beginning in
158 2008/2009, NRSA surveys ~2,000 stream/river locations that are selected using a probabilistic
159 survey design and handpicked by state and tribal partners. These sites are surveyed for
160 biological, chemical, and physical characteristics using standardized field protocols (USEPA
161 2023). Only sites included in the probabilistic sample are selected to be representative of the
162 entire population of streams and rivers in the contiguous US and used for assessment (Olsen and
163 Peck 2008). In addition, approximately 10% of the probabilistic sites are revisited to assess
164 within-year variability and approximately 30% of these sites are resampled in the next survey
165 cycle to evaluate among-year variability.

166

167 *Benthic macroinvertebrate assemblages and physiochemical variables*

168 Procedures for collecting benthic macroinvertebrates are described in detail elsewhere
169 (Hughes and Peck 2008, USEPA 2017a). Briefly, each survey location was defined as a length of
170 stream or river equal to a multiple of its channel width. In wadeable and boatable sites, the reach
171 length sampled was equal to 40 channel widths, a minimum of 150 m, or a maximum of 4km.
172 Macroinvertebrate samples were collected along 11 equally spaced cross-section transects along
173 the reach. In wadeable sites, samples were collected in an alternating left, center, right order
174 along the transects using a D-frame kick net (500-um mesh, 0.09 m² area). In boatable sites,
175 samples were collected along the left or right wadeable margin from a 1m linear sweep of the
176 dominant habitat using a D-frame kick net (500-um mesh, 0.3048 m² area). Samples from each
177 survey location were combined into a single composite sample, preserved in ethanol, and sent to
178 a taxonomist for subsampling and identification.

179 To assess the effects of anthropogenic disturbance on in-stream physiochemical
180 conditions and in turn macroinvertebrate assemblages, we focused on in situ nutrient
181 concentrations, salinity, and physical habitat variables. These environmental gradients are
182 commonly altered by anthropogenic activities in the surrounding watershed and frequently used
183 as abiotic screens to identify least-disturbed reference sites (Herlihy et al. 2008, Paulsen et al.
184 2008, USEPA 2023). Specifically, we selected total nitrogen (NTL) and total phosphorus (PTL)
185 as indicators of excess nutrients supplied from agriculture and/or urbanization (Herlihy et al.
186 1998) and chloride (CL) and sulfate (SO₄) as indicators of salinity. Elevated CL is a general
187 indicator of human disturbance in the catchment (Herlihy et al. 1998) and SO₄ can indicate sites
188 that are affected by mine drainage (Herlihy et al. 1991). Mean substrate diameter (SUBD) and

189 riparian disturbance index (RPDI) were selected as indicators of physical habitat because human
190 activities can increase fine sediment inputs or directly modify the riparian area (Kaufmann 1999,
191 Kaufmann et al. 2022b). In addition, we included mean summer air temperature (Mean July and
192 August temperature, MSAT) and total annual precipitation (TPRCP) to assess potential changes
193 associated with differences from a 1900-1950 baseline. Although other factors are affected by
194 human disturbances (e.g., metals and pesticides and hydrologic alteration), these variables
195 provide a reasonably broad characterization of the physiochemical environment at a site (Herlihy
196 et al. 2008, Paulsen et al. 2008)

197 During the survey, a single water sample was collected at each site and shipped to a
198 central analytical laboratory. PTL and NTL were measured by persulfate digestion and
199 colorimetry and CL and SO₄ were measured by ion chromatography (USEPA 2017b). SUBD is
200 the geometric mean of the numeric value assigned to substrate size classes measured in the field
201 (Kaufmann 1999). RPDI summarizes the presence/absence of 11 categories of human
202 disturbance, including buildings, landfill/ trash, logging, mining, developed parks or lawns,
203 pavement or cleared lots, pipes (withdrawal or wastewater), roads, row crops, pastures or
204 hayfields, and walls or revetments in the riparian area, adjacent to each to the 11 transects where
205 macroinvertebrates were collected (Kaufmann 1999, USEPA 2017a, Kaufmann et al. 2022b). In
206 addition, we included mean summer air temperature (Mean July and August temperature,
207 MSAT) and total annual precipitation (TPRCP) for either 2018 or 2019 (depending on sample
208 year) obtained from the PRISM Climate Group (<https://prism.oregonstate.edu/>).

209

210 *Modeling genus-environment relationships*

211 We quantified relationships between macroinvertebrate assemblages and eight
212 physiochemical variables (i.e. NTL, PTL, CL, SO₄, SUBD, RPDI, MSAT, and TPRCP) using
213 joint species distribution models (JSDMs) fitted with the Hierarchical Modeling of Species
214 Communities R package (Ovaskainen and Abrego 2020, Tikhonov et al. 2020). JSDMs are a
215 multivariate hierachal generalized linear mixed model fitted with Bayesian inference. They are
216 uniquely suited to evaluate relationships between anthropogenic disturbance and biological
217 assemblages because they are multi-species models that quantify taxon-environment
218 relationships for all members in an assemblage simultaneously and account for unmeasured
219 variables, including abiotic factors and species associations, using random effects specified at the
220 sample-level (Warton et al. 2015, Ovaskainen and Abrego 2020, Deflem et al. 2021). Indeed,
221 sample-level random effects are meaningful for multivariate models because they are not
222 confounded by residual variation as with univariate models. More specifically, this attribute
223 allows models to account for nonindependence among residuals for each site and improves
224 estimates of the fixed effects (i.e. taxon-environment relationships) (Ovaskainen et al. 2016). In
225 addition, JSDMs also allow for the inclusion of phylogenetic relatedness as a hierarchical term
226 that can lend additional improvements to estimated taxon-environment relationships.

227 We provide a detailed description of the modeling framework and application elsewhere
228 (Kopp et al. 2023). In brief, we used presence/absence data from 1,891 benthic
229 macroinvertebrate assemblages surveyed as part of the probabilistic and handpicked sites and
230 focused on taxa that were collected from a single site visit, identified to genus, and occurred at
231 ≥10% of sites within an ecoregion (Table 1). Separate models were fit for each ecoregion
232 boundaries to define regional species pools and thus assume that environmental conditions are
233 the primary factor driving genus occurrence (Chessman and Royal 2004). All physiochemical

234 variables were measured in the field during the survey, except for MSAT and TPRCP, which
235 were obtained from PRISM Climate and matched to the appropriate survey year (Table 1).
236 Physicochemical variables were used as linear fixed effects and sample-level random effects were
237 used to statistically control for unmeasured variables. Previously, we found few genus-
238 environment relationships were unimodal (Kopp et al. 2023) and therefore assumed linear
239 relationships were appropriate for this study. Our analysis focused on relatively small number of
240 environmental variables because these are commonly altered by human activities. We also use
241 taxonomy as a surrogate for phylogenetic relatedness as a hierarchical term in the model. All
242 models were fitted with the default prior distributions (Ovaskainen and Abrego 2020), using
243 three independent chains (3,000 posterior samples). Convergence was determined to be
244 satisfactory by potential scale reduction factor < 1.1 for fixed effects and phylogenetic parameters.

245 Our primary motivation for using these models was to measure how macroinvertebrate
246 assemblages may change if the influence of anthropogenic disturbances were removed from in
247 stream physicochemical variables and all else remained unchanged (Figure 2). Since we do not
248 use these models to predict to new locations, model performance was primarily evaluated with
249 respect to its explanatory power, i.e. the fitted models ability to reproduce the observed genus
250 richness and composition (Wilkinson et al. 2021, Abrego and Ovaskainen 2023). Predicted,
251 taxon-specific occurrence probabilities were summed to obtain predicted richness and regressed
252 against the observed richness (Calabrese et al. 2014). We determined model acceptability based
253 on three criteria: $R^2 \geq 0.2$, $-1.5 \leq \text{intercept} \leq 1.5$, and $0.85 \leq \text{slope} \leq 1.15$ (Linke et al. 2005).
254 We calculated model performance metrics using each of 3,000 posterior samples to obtain a
255 distribution of plausible estimates of performance metrics and report the mean and 5th and 95th
256 quantiles as measures of uncertainty. In addition, we assessed compositional similarity between

257 predicted and observed assemblages using a probabilistic adaptation of Jaccard similarity
258 (Scherrer et al. 2020) to avoid introducing error associated with converting predicted occurrence
259 probabilities into binary outcomes (Calabrese et al. 2014). We calculated similarity for each site
260 using the mean predicted occurrence probabilities and report the mean and 5th and 95th quantiles
261 across all sites.

262 Predictive power for models that use random effects can only be assessed for cases where
263 at least some sampling units were included in the calibration data (Abrego and Ovaskainen
264 2023). As part of NRSA, a subsample of sites are revisited with the intent of assessing temporal
265 variability in metrics and indices (Table 1). Since we fit the JSMDs using sample-level random
266 effects, these data were used to as a second type of validation. Specifically, we evaluated
267 whether the richness observed at a revisited site was within the posterior distribution of the fitted
268 models. Although this metric of validation may seem less restrictive compared to those used to
269 evaluate explanatory power, we expected variation between samples to be rather large because of
270 stochastic events (e.g. high or low stream flows), ecological processes (e.g. emergence and
271 dispersal), and sampling procedures (e.g. field collection and laboratory subsampling) that our
272 models were not designed to capture. Given this level of potential uncertainty, this metric was
273 intended to address whether an observation could have come from the same process that our
274 model was intended to capture. We also compared the assemblages collected during the revisit to
275 the predicted assemblages using the probabilistic adaptation of Jaccard similarity (Scherrer et al.
276 2020) and report the mean and 5th and 95th quantiles across sites.

277

278 *Modeling physiochemical gradients*

279 We used random forest models to relate anthropogenic and landscape geoclimatic factors
280 to total nitrogen (NTL), total phosphorous (PTL), chloride (CL), sulfate (SO₄), and substrate
281 diameter (SUBD) and make predictions if anthropogenic disturbances were removed. The suite
282 of predictor variables were selected based on their hypothesized relationship with stream water
283 chemistry and bed particle size (Lin et al. 2021, Zak et al. 2021, Kaufmann et al. 2022b, Kaushal
284 et al. 2023, Sabo et al. 2023). Geoclimatic factors included watershed morphology (e.g., basin
285 area, elevation, and slope) and lithological characteristics (e.g., lithological phosphate, nitrogen,
286 and sulfur, and soil erodibility). Anthropogenic factors included percent agriculture, road
287 density, and presence of mines and impoundments (Appendix S1: Table S1). Predictor variables
288 were obtained or derived from the StreamCat Database (Hill et al. 2016), National Atmospheric
289 Deposition Program (nadp.slh.wisc.edu), and EPA's National Nutrient Inventory (Sabo et al.
290 2019, Lin et al. 2021, Sabo et al. 2021, Sabo et al. 2023).

291 Separate random forest models were fit for NTL, PTL, CL, SO₄, or SUBD using the
292 entire national dataset to maximize the range of the predictor variables included in each model.
293 Prior to model fitting the dataset was randomly split into training and testing portions (80% and
294 20%, respectively). Model fit was assessed by the coefficient of determination (R²) and root
295 mean squared error (RMSE) on training and testing portions of the dataset. The relative
296 importance of the covariates was evaluated by the change in mean squared error after
297 permutating each variable (%IncMSE). Because variable importance can be influenced by
298 correlated variables, we confirmed that all variables used in the models had pairwise Pearson's
299 correlation coefficients < 0.7 and the variance inflation factors (VIF) were between 2.6 and 3.9.
300 (Appendix S1: Table S3) In general, VIF > 5 indicates a potential problem with multicollinearity
301 (O'brien 2007). Partial dependence plots were used to visualize the relationship between the

302 most important anthropogenic factors and each physiochemical variable. Random forest analysis
303 was performed using the quantregForest R Package (Meinshausen 2017).

304

305 *Evaluating effects of human disturbance on physiochemical gradients*

306 We used the fitted random forest models to predict values for NTL, PTL, CL, SO4, and
307 SUBD if anthropogenic disturbance was removed by setting all anthropogenic factors to the
308 lowest value observed in our dataset (often zero, Appendix S1: Table S2) and leaving
309 geoclimatic factors unchanged. Hereafter “hindcast” refers to the removal of anthropogenic
310 disturbance from the physiochemical variables. Although we included the entire range of
311 predictor values, it is possible that hindcast data are not sufficiently similar to the data used to
312 train the model and thus susceptible to extrapolation (Meyer and Pebesma 2021, Yuan et al.
313 2024). To test whether the hindcast dataset was sufficiently similar to the data used to train the
314 models, we first mean-centered and scaled all predictor variables to equivalent units (i.e.
315 standard deviations) and weighted them according to their importance in the model. We then
316 calculated the minimum Euclidean distance between each site in the hindcasted data, and each
317 site used in the training dataset using the same center and scale. The minimum value was then
318 divided by the mean Euclidean distance among all training data. Following Yuan et al. (2024),
319 we flagged hindcast predictions for sites that exceeded 0.5 as susceptible to extrapolation.

320 We evaluated the effects of anthropogenic disturbance on each physiochemical variable
321 based on deviations from the hindcast estimates using standardized anomalies (i.e. z-scores), or
322 the difference between the observed and hindcast value scaled by standard deviation. For
323 variables we modeled using random forests, we standardized the difference between observed
324 and hindcasted conditions using the RMSE of each model (Kilgour and Stanfield 2005). For

325 climate variables, we obtained baseline historical climatic values (MSAT and TPRCP) as the
326 mean summer air temperature for 1900-1950 from PRISM climate data
327 (<https://prism.oregonstate.edu/historical/>). We then standardized the difference between the
328 present-day values (i.e. 2018 or 2019) and the baseline using the standard deviation of the 50yr
329 dataset. Importantly, using standardized anomalies rather than absolute differences has
330 advantages because it accounts for unexplained variation in the model or natural variability
331 among sites (Kilgour and Stanfield 2005). Furthermore, because the values are in units of
332 standard deviations, thresholds to evaluate whether the magnitude of difference is sufficient to
333 support an effect of anthropogenic disturbance can be rather intuitive. Specifically, we expected
334 anthropogenic disturbance to elevate NTL, PTL, CL, and SO4 concentrations and increase or
335 decrease SUBD, MSAT, and TPRCP. Accordingly, we identified sites with observed
336 concentrations $>2SD$ from the hindcast values for NTL, PTL, CL, and SO4 and $> |2SD|$ for
337 SUBD, MSAT, and TPRCP as having evidence of anthropogenic disturbance.

338 Since RPDI is a direct measure of anthropogenic disturbance, we initially set this variable
339 to zero for all locations but found that this value may be too strict because of the number and
340 diversity of factors included in the index and did not provide much insight into potential regional
341 variation in disturbance. Instead, we used 0.33 as a threshold which is interpreted as one type of
342 human disturbance observed within 10 m of the streambanks at no more than one third of the 22
343 riparian plots sampled, on average (Kaufmann et al. 2022b, USEPA 2023). Although this does
344 mean that not all sites were completely free of riparian disturbance, it is consistent with other
345 studies that have used this index to evaluate human disturbance in the riparian area (Kaufmann et
346 al. 2022a, Kaufmann et al. 2022b, USEPA 2023).

347 For each physiochemical variable we estimated the total percent of streams that were
348 potentially affected by anthropogenic disturbance (i.e. standardized anomaly > 2SD) using only
349 the probabilistic samples and their weights reflective of the entire population of streams and
350 rivers assigned to them by NRSA (USEPA 2023). Estimates for each ecoregion were generated
351 using the cat_analysis() function from the spsurvey R package (Dumelle et al. 2023). In
352 addition, we tallied the number of physiochemical variables that were potentially affected at each
353 site to elucidate instances where human disturbance affects multiple environmental variables
354 simultaneously.

355

356 *Evaluating effects of human disturbance on macroinvertebrate assemblages*

357 We evaluated the effects of removing anthropogenic disturbance from the
358 physiochemical environment on macroinvertebrate assemblages using the fitted JSMDs. For the
359 sites that had evidence of human disturbance (i.e. standardized anomaly > 2SD), we substituted
360 the hindcasted value (either predicted from random forest model, 1900-1950 averages for MSAT
361 and TPRCP or 0.33 for RPDI) for the present day value in the dataset and used these data to
362 predict macroinvertebrate assemblages that could occur if disturbance was removed or reduced
363 (Chessman and Royal 2004). To evaluate the relative effects of hindcasting each physiochemical
364 variable, we created 4 scenarios changing either NTL and PTL (Nutrient Scenario), CL and SO₄
365 (Salinity Scenario), RPDI and SUBD (Habitat Scenario), MSAT and TPRCP (Climate Scenario)
366 and leaving the others at their observed values. In addition, we predicted the macroinvertebrate
367 assemblage after changing all physiochemical variables to their hindcast value (Hindcast
368 Scenario). For each scenario we compared predicted genus richness from present-day conditions
369 to hindcast genus richness.

370 JSDM predictions are three-dimensional data arrays that contain 3,000 posterior samples
371 of site-specific occurrence probabilities for each genus (i.e., site x genera x posterior samples).
372 For each posterior sample we summed predicted occurrence probabilities for all genera to obtain
373 3,000 plausible estimates of hindcast genus richness for each site. We then compared the mean
374 present-day genus richness (i.e., predicted from JSDM using present-day values for the
375 physiochemical variables) to the hindcast posterior distribution. We evaluated whether present-
376 day richness was below the 10th or 25th quantiles, indicating a reduction in present-day genus
377 richness relative to hindcast predictions, or above the 75th or 90th quantiles, indicating an increase
378 in genus richness relative to hindcasted (Figure 2). Hereafter, present-day values that are located
379 in the extreme ends of the distribution are described as having either >0.75 or >0.90 support for a
380 difference from hindcasted predictions, respectively. We assessed the effects of removing
381 anthropogenic disturbance from each of the 4 combinations of environmental variables (i.e.
382 Nutrient Scenario, Salinity Scenario, Habitat Scenario and Climate Scenario) and the effects of
383 removing anthropogenic disturbance from all environmental variables simultaneously (Pristine
384 Scenario) with >0.75 and >0.90 support as evidence for an effect of disturbance on
385 physiochemical variables.

386 Identifying genera having site-specific occurrence probabilities that differ between
387 present-day and hindcast conditions helps interpret assemblage-level changes in response to
388 anthropogenic disturbance. We identified genera that had higher occurrence probability under
389 present day conditions as “increasers” and genera with lower occurrence probability as
390 “decreasers”. Differences were determined with >90% support. For each genus we calculated the
391 proportion of sites where they increased or decreased. If a genus was identified as an increaser at
392 a large proportion of sites it could become more prevalent in response to anthropogenic

393 disturbance. Alternatively, if a genus was identified as a decreaser at a large proportion of sites,
394 it could become less prevalent in response to human disturbances. For each ecoregion, we report
395 the mean for major taxonomic groups (e.g. insects and non-insect genera) to further understand
396 taxon-specific trends in the context of biodiversity loss (Jähnig et al. 2021, Rumschlag et al.
397 2023).

398 Identifying increaser and decreaser genera could also enhance our ability to measure the
399 effects of removing anthropogenic disturbance because assemblage composition could be
400 affected by anthropogenic disturbance without a subsequent change in richness (Van Sickle
401 2008). To assess compositional differences between present-day and hindcast assemblages we
402 used increaser/decreaser assignments to create two community matrices, one representing
403 present-day assemblage and the other representing hindcast assemblage. Increasers were
404 assigned a value of 1 in the present-day matrix and 0 in the hindcast matrix because they had a
405 higher occurrence probability under present-day conditions. Conversely, decreasers were
406 assigned 0 in the present-day matrix and 1 in the hindcast matrix because they had a significantly
407 higher occurrence probability under hindcasted conditions. For genera that did not have
408 sufficient support (< 0.90) for changing occurrence probabilities were considered not to be
409 affected by anthropogenic disturbance and were assigned 1 in both matrices. We compared the
410 two matrices (i.e. site \times genera) using Jaccard similarity index and identified sites with a
411 similarity score of < 0.9 as changing compositionally.

412 We assessed the consequences of removing anthropogenic disturbance from in-stream
413 physiochemical environment for macroinvertebrate assemblages by identifying sites within each
414 region that had evidence for either a change in richness (>0.75 support) or composition (Jaccard
415 Similarity < 0.9) from present-day conditions. Then, using only the probabilistic samples that

416 were not flagged for extrapolation and had complete data ($n = 1748$) and their weights, estimated
417 the total percent of streams within each ecoregion that were potentially affected by
418 anthropogenic disturbance. Estimates for each ecoregion were generated using the cat_analysis()
419 function from the spsurvey R package (Dumelle et al. 2023). In addition, we report results with
420 different levels of support (> 0.75 and > 0.90) for richness change and separate compositional
421 change to convey uncertainty and methodological differences.

422

423 Results:

424 *Random Forest Modeling*

425 The random forest models explained 46 to 77% of the variation in the training data
426 ($n=1502$) and 51 to 78% of the variation in the test/validation data ($n=375$). The model of
427 substrate diameter (SUBD) had the highest RMSE in both training and testing datasets compared
428 to models of total nitrogen (NTL), total phosphorous (PTL), chloride (CL), and sulfate (SO₄)
429 (Table 2). Variable importance for each physiochemical model, measured by the percent change
430 in MSE after permutation, revealed that runoff was among the most important geoclimatic
431 variables for each model and that agricultural landcover in the watershed or riparian area ranked
432 among the most important anthropogenic variables (Appendix S1: Table S2). Percent agricultural
433 landcover in the watershed was the most important variable predicting NTL and PTL, the second
434 most important for predicting SO₄ and third most important for predicting CL. Other
435 anthropogenic variables were also ranked relatively high in importance. For example, road
436 density in the watershed was the second most important variable predicting CL and fifth most
437 important variable for SO₄, while percent natural vegetation cover in the riparian area was the
438 third most important variable predicting SUBD.

439 We visualized the effects of the anthropogenic factors using partial dependence plots
440 (Figure 3). In general, the anthropogenic factors were associated with increased nutrients and
441 salinity and decreased substrate diameter. For example, percent crops and anthropogenic nutrient
442 inputs of nitrogen and phosphorous were positively associated with NTL and PTL. Road density
443 and mean summer air temperature had a positive association with CL while total annual
444 precipitation had a weakly negative relationship. Density of coal mines in the watershed had a
445 relatively strong positive effect on SO₄ and SUBD had a strong positive relationship with natural
446 vegetation cover and strong negative relationship with agricultural activities in the riparian area.

447

448 *JSDM assemblage level performance*

449 We fitted JSDMs for each region using presence/absence data for 59 – 127 genera from
450 152 – 266 sites surveyed during the 2018 – 2019 NRSA cycle (Table 1). The total number of
451 sites used to fit JSDMs were higher (1,891) than the sites we ultimately assessed because we
452 included probabilistic and handpicked sites (USEPA 2023). We found that summing predicted
453 occurrence probabilities for all genera modeled could potentially overestimate observed genus
454 richness (i.e. observed vs predicted richness intercepts < -1.5). Upon further inspection we found
455 that this overprediction was due to many genera having exceptionally low predicted occurrence
456 probabilities but nonetheless present in the regional species pool. Thus, we used a threshold to
457 exclude genera with predicted occurrence probabilities < 0.05 or < 0.10 prior to summation to
458 correct this bias (Table 3).

459 Because site level random effects accounted for unmeasured factors in the models, we did
460 not expect our models to have strong predictive power when projected to new locations (Abrego
461 and Ovaskainen 2023, Kopp et al. 2023). However, we could assess whether values obtained at

462 revisited sites were within the predicted posterior distribution of the fitted models as a secondary
463 form of validation. Although the number of revisited sites available for model testing was low
464 (Table 1), we found that >90% of the genus richness values observed during revisit sampling
465 were within the posterior distribution of the models. Importantly, this lends plausibility that the
466 posterior distribution reflects processes characterizing macroinvertebrate assemblages (Table 3
467 and Appendix S1: Figure S2).

468 The probabilistic adaptation of Jaccard similarity revealed that the predicted and
469 observed assemblages were on average >50 similar (Table 3). Sites located the WMT tended to
470 have the most similar assemblages (mean = 0.6, 5th quantile = 0.36, 95th quantile= 0.75) while
471 sites in the XER tended to have the least similarity (mean = 0.23, 5th quantile = 0.23, 95th
472 quantile= 0.71). Although potentially vulnerable to sample size constraints, mean similarity
473 between model predictions and assemblages observed at revisited sites were also relatively high,
474 with mean values ranging between 0.35 and 0.68 (Table 3).

475

476 *Disturbance effects on physiochemical conditions*

477 We tested whether using the random forest models to predict physiochemical conditions
478 after removing the effects of anthropogenic activities were susceptible to extrapolation. We
479 found predictions at 77 of the sites (~4%) were potentially susceptible to extrapolation. This
480 indicates that removing disturbance would cause some sites to be sufficiently different from the
481 data used to train the models and suggest that these sites do not have a natural analog. Indeed,
482 most of the sites flagged for potential extrapolation were located in the Coastal Plains Ecoregions
483 (18%) and were associated with models used to predict NTL concentrations in the lower
484 Mississippi and SUBD along the southern coast (Appendix S1: Figure S1). To avoid the

485 potential for bias associated with extrapolation, we designated them as “Not Assessed” in our
486 analysis.

487 For sites that were not potentially susceptible to extrapolation, removing anthropogenic
488 disturbance from the physiochemical environment generally decreased nutrient concentrations,
489 salinity, and riparian disturbance and increased substrate diameter (Figure 4). Change in NTL
490 and PTL concentrations were most evident in Temperate Plains and Southern Plains and change
491 in CL was most evident in the Northern Appalachians. Substrate diameter was noticeably coarser
492 after hindcasting for all regions except Xeric, Southern Appalachians, and Northern
493 Appalachians. Observed and hindcasted values for sulfate, and mean summer air temperature
494 were generally similar for all ecoregions.

495 Anthropogenic disturbance affected physiochemical gradients differently depending on
496 the ecoregion (Figure 5a). We found 60.8% (95% Confidence Interval, hereafter CI = 52.8-
497 68.8%) of streams in Temperate Plains and 38.4% (CI = 32.0-44.9%) of streams Upper Midwest
498 had elevated NTL relative to hindcasted estimates. In the Northern Appalachians, 41.4% (CI =
499 34.5-48.2%) of streams were found to have potentially elevated CL concentrations. Streams with
500 excess fine substrates were most prevalent in the Temperate Plains (29.2%, CI = 21.2-37.1%)
501 and Northern Plains (21.4%. CI = 14.1-28.9%). We also found similarities among regions. RPDI
502 and TPRCP were the most prevalent disturbance among all ecoregions while disturbed SO4 and
503 MSAT values were the least prevalent. Finally, anthropogenic disturbance could alter multiple
504 physiochemical variables simultaneously (Figure 5b). In general sites with >2 physiochemical
505 variables disturbed by anthropogenic activities were located in Northern Appalachians, Southern
506 Plains, Temperate Plains, and Upper Midwest.

507

508 *Macroinvertebrate assemblage response to undisturbed physiochemical environment*

509 Macroinvertebrate models were fit for each region separately and we assessed whether
510 removing anthropogenic disturbance would produce estimates that exceed the range of values
511 used to fit each model. Hindcast predictions exceeded the range of present-day conditions at 1
512 location in CPL, SAP and XER, 6 locations in TPL, and 12 locations in WMT. In these
513 instances, we reset the hindcasted values to be within the range (minimum) of data used to fit the
514 mode to avoid potential extrapolation.

515 Hindcast predictions for each category of physiochemical variables generated different
516 outcomes for macroinvertebrate assemblages (Figure 6a). For example, removing anthropogenic
517 disturbance from NTL and PTL concentrations, produced macroinvertebrate assemblages with
518 higher genus richness in the Temperate Plains and lower richness at some sites in the Southern
519 Plains and Southern Appalachians ecoregions (Figure 6a). On the other hand, removing
520 disturbance from CL and SO₄ tended to increase genus richness in the Southern Plains but
521 decreased richness at sites in the Northern Appalachians. Hindcasted physical habitat variables,
522 mostly affected macroinvertebrate assemblages in the Northern Plains by decreasing richness but
523 also increased genus richness in Coastal Plains, and Northern Appalachians and Upper Midwest
524 (Figure 6a). TPRCP generally contributed to changes in macroinvertebrate assemblages because
525 relatively few sites had MSAT temperature anomalies exceed 2 standard deviations from 1900-
526 1950 mean. In the Southern Appalachians genus richness decreased from present day conditions
527 while in the Northern Plains genus richness generally increased from present day conditions.
528 Removing hindcasting all physiochemical variables typically increased genus richness in all
529 regions except for the Northern Plains and Southern Appalachians (Figure 6b).

530 Identifying genera that had higher present-day occurrence probabilities (increasers), or
531 higher hindcast occurrence probabilities (decreasers) indicated that insects tended to be
532 decreasers while non-insects tended to be increasers, but this was not consistent across all
533 regions (Table 4). For example, in the Southern Plains and Western Mountains, insects were
534 almost equally likely to be increasers (0.1 and 0.6, respectively) or decreasers (0.09 and 0.03,
535 respectively) and in the Northern Plains both insects and non-insects tended to be increasers at a
536 larger proportion of sites (0.19 and 0.22). Among insects, members of chironomidae were
537 generally most likely to be increasers. Among non-insects, in contrast, members of mollusca
538 were most likely increasers (Table 4). Thus, higher present-day genus richness compared to
539 hindcast estimates could be due to genera that are typically considered to be tolerant of human
540 activities.

541 Using the probabilistic sites, we aggregated our results to reflect the total population of
542 streams and rivers in each ecoregion and detected that 14.3 – 75.5% of streams have present-day
543 assemblages that differ from assemblages expected from the hindcasted physiochemical
544 environment (Figure 7). With respect to changes in richness with support >0.75, the Northern
545 Plains (58.3%), Temperate Plains (54.8%), and Southern Plains (54.2%) the most streams with
546 potentially altered macroinvertebrate assemblages while the Western Mountains (9.2%), Upper
547 Midwest (19.6%), and Northern Appalachians (29.7%) have the fewest. Compositional change
548 without a corresponding change in genus richness could increase the percentage of streams with
549 altered assemblages. Evidence for compositional change was most pronounced in Temperate
550 Plains (20.7%), Southern Appalachians (20.7 %) and Upper Midwest (18.8%). All regions had a
551 percentage of streams that could not be assessed because of potential extrapolation of hindcasted
552 physiochemical variables, incomplete data or macroinvertebrate assemblages consisting of only

553 rare genera (i.e. prevalence < 10%). Among them, the Coastal Plains (12.9%) and Xeric (11.1%)
554 had the largest percentage of streams that were not assessed, thus the total number of sites
555 assessed was less than the total number of probabilistic sites surveyed (n = 1748).

556

557 Discussion

558 Assess the effects of anthropogenic disturbance on biological assemblages is challenging
559 because few minimally disturbed sites remain (Stoddard et al. 2006). Sites in least-disturbed are
560 often adopted as an alternative but are difficult to consistently define, vary in quality, and can be
561 spatially aggregated (Hawkins et al. 2010, McNellie et al. 2020). The model-based framework
562 we developed contributes to a suite of other approaches intended to circumvent the need for
563 reference sites (Chessman and Royal 2004, Elias et al. 2016, Yuan et al. 2024). Specifically, we
564 evaluated the effects of human disturbance on several physiochemical variables and, in turn,
565 addressed whether altered physiochemical conditions affect macroinvertebrate assemblages.

566 Using this approach, we found that anthropogenic disturbance can affect multiple
567 physiochemical variables simultaneously and that the effects on any single factor can vary
568 among ecoregions. We also found that removing or reducing disturbance could change genus
569 richness at >50% of the streams in some ecoregions and up to 75% if compositional change is
570 considered. Collectively, our framework offers a promising alternative to evaluating the effects
571 of specific disturbances on macroinvertebrate assemblages that does not rely on reference sites.

572

573 *Hindcasting physiochemical conditions*

574 Random forest models have been used by others to predict physiochemical variables if
575 human disturbance was reduced or removed (Yuan et al. 2024). In general, using models to infer

576 physiochemical conditions eliminates the need to identify reference sites or collect excessive
577 data from undisturbed locations (Herlihy and Sifneos 2008, Soranno et al. 2011). In this study,
578 random forest models improve on earlier regression-based hindcasting approaches (Dodds and
579 Oakes 2004, Herlihy and Sifneos 2008) because they incorporate a relatively large number of
580 natural and anthropogenic predictor variables and accommodate their complex relationships with
581 the response variables (Yuan et al. 2024). Importantly, because reference sites may not be
582 representative of all sites that need to be assessed including natural and anthropogenic variables
583 enabled us to make site-specific predictions rather than relying on an estimated mean value (i.e.
584 the intercept from a multiple regression as a function of only disturbance) for an entire region
585 (Dodds and Oakes 2004, Herlihy and Sifneos 2008). Thus, our hindcasting approach provides a
586 first-order approximation for what conditions at a site could be if they are presently unknown.

587 Furthermore, random forest models permitted us to evaluate the relative importance of
588 specific anthropogenic activities and characterize potential site and regional scale differences
589 after reducing or removing them. For example, higher than expected total nitrogen in Temperate
590 Plains and Upper Midwest and the relatively high importance of percent agriculture landcover
591 our model suggests that nutrient inputs associated with agricultural activities may be elevating
592 in-stream nutrient concentrations (Lin et al. 2021, Sabo et al. 2023). Similarly, elevated chloride
593 concentrations in the Northern Appalachians and the relative high importance of road density in
594 the model, could reflect the contribution of road salts to freshwater salinization (Kaushal et al.
595 2023).

596 Certainly, inferring physiochemical conditions for different levels of human disturbance
597 depends on the quality and structure of the model. Although our models were evaluated using an
598 independent validation dataset, hindcasting required a dataset that reduced or eliminated the

599 effect of predictor variables associated with human disturbance. As consequence, the dataset
600 used for hindcasting could be dissimilar from the data used to calibrate the model and potentially
601 generate errors associated with extrapolation (Meyer and Pebesma 2021). For most regions, the
602 hindcasting dataset was sufficiently similar to the training dataset such that we could assess the
603 alteration of physiochemical variables for majority of sites. Nonetheless 18% of sites surveyed
604 the Coastal Plains were flagged as a result of testing for extrapolation. Considering that many of
605 the sites were located in the lower Mississippi River Basin, a unique and heavily modified
606 system, it is perhaps unsurprising that removing human disturbances from this system is beyond
607 the domain of our model. Testing for potential extrapolation in the context of hindcasting is
608 novel and, although the threshold to determine whether our results were prone to extrapolation
609 has been used previously (Yuan et al. 2024), more research should be devoted to optimizing and
610 interpreting this threshold. When locations or streams are potentially susceptible to extrapolation,
611 it might be more reasonable to select values that represent management goals, best professional
612 judgement or those that maximize societal benefits (Bouleau and Pont 2015).

613 To assess how human-related disturbance effects physiochemical environment, we used
614 standardized anomalies (i.e. z-scores), to account for either unexplained variation in the model
615 values or variation along a baseline. Thresholds are of central importance for the communication
616 and evaluation of human disturbances communication (Wood 2008). Here, we used a $\pm 2\text{SD}$
617 threshold for deciphering whether human disturbances alter physiochemical conditions (Kilgour
618 and Stanfield 2005). Although this threshold is intuitive, it could potentially be too strict for
619 some variables. For example, values for MSAT at many locations did not exceed the threshold in
620 our analysis, suggesting that 2018 and 2019 temperature averages are consistent with 1900-1950
621 averages, given interannual variation. However, warming of 1SD has been implicated in

622 reductions of insects in agricultural landscapes (Outhwaite et al. 2022) and there is evidence for
623 increasing summer air temperatures in the United States of approximately 0.09°C per decade
624 since 1901 (USEPA 2024a). On the other hand, to evaluate riparian disturbance, we used a
625 threshold of 0.33 which suggests that human disturbance is not completely absent for all sites
626 that were below this threshold. Indeed, nearly every site in our analysis has RPDI >0 such that
627 this threshold did not informatively differentiate regional variation in disturbance. Further, this is
628 the threshold used by others to identify least disturbed sites (Kaufmann et al. 2022b, USEPA
629 2023) and we sought to evaluate how this could affect microinvertebrate assemblages. Although
630 thresholds were necessary to identify which regions may have relatively more disturbed
631 physiochemical conditions that others, future applications may select different thresholds, report
632 continuous scores, or vary the amount of disturbance to meet specific applications (Yuan et al.
633 2024).

634

635 *Application for biological assessment without reference sites*

636 Using hindcast physiochemical variables to predict the assemblage expected to occur if
637 anthropogenic disturbance was removed allowed us to infer the consequences of altering a
638 relatively small number of physiochemical variables. Central to our application is that
639 multivariate models can implement meaningful site-level random effects that statistically control
640 for unmeasured environmental variables and potential biotic interactions (Warton et al. 2015,
641 Ovaskainen et al. 2016, Kopp et al. 2023). Although, this feature kept our model sufficiently
642 tractable it limits our ability to make accurate predictions to new locations (Abrego and
643 Ovaskainen 2023). However, predicting to new locations was not our objective. Rather, our
644 models establish empirical relationships between macroinvertebrate occurrences and then

645 evaluate how those assemblages might differ if select physiochemical conditions changed while
646 all else remained constant. Furthermore, because the data we used for our analysis was collected
647 as part of a probabilistic survey, designed to be representative of the population of stream and
648 rivers (Olsen and Peck 2008), our site-specific inferences can be aggregated to elucidate regional
649 and sub-continental patterns. Indeed, it may be undesirable to refit models every time an
650 assessment is needed, and future efforts should focus on improving predictive abilities by
651 including immutable factors in addition to stressor gradients (Yuan et al. 2024).

652 Importantly our approach differs significantly from traditional, reference site-based
653 approaches (Hawkins et al. 2000, Herlihy et al. 2008, USEPA 2023). Foremost, our efforts focus
654 explicitly on a relatively small number of potential stressors whereas reference-site approaches
655 focus implicitly on a theoretically larger, but undefined number of stressors. For example, the
656 difference between a test site and reference sites could be related to a number of other stressors
657 that were not used as biotic screens but co-occur with them (Herlihy et al. 2008, Paulsen et al.
658 2008). Alternatively, in our model-based approach, differences between present day and hindcast
659 assemblages are only related to changes in the physiochemical variables included as fixed
660 effects. Although the former may elucidate general disturbance effects, it is difficult to attribute
661 differences between assemblages to a specific environmental disturbance (Paulsen et al. 2008).
662 On the other hand, our model-based approach enhances interpretations with respect to specific
663 physiochemical variables but may omit important anthropogenic stressors that were not
664 explicitly included in the model. Thus, the choice between reference site-based and model-based
665 assessments may be contingent on whether the environmental gradients that are commonly
666 disturbed by human activities can be identified and appropriately modeled.

667 The JSDMs were also fitted using bayesian inference and, as such, yielded posterior
668 distributions that can be used for hypothesis testing (Johnson et al. 2022). Specifically, we
669 evaluate whether the effects of human disturbance on physiochemical variables were sufficient to
670 alter benthic assemblages on a site-specific basis. In contrast, traditional reference site-based
671 approaches evaluate each test site based on quantiles of the distribution obtained from reference
672 sites (Herlihy et al. 2008, USEPA 2023). In this regard, JSDMs and bayesian inference may be
673 advantageous in interpreting biological condition estimates because they account for
674 uncertainty in predictions for each site. Furthermore, we also included all genera from the
675 regional taxa pool while some typical reference site-based approaches include only taxa that
676 occur at reference sites. This implies that in the absence of human disturbance those taxa should
677 occur at all locations regardless of other taxa. This is problematic because genera that tolerate
678 anthropogenic disturbance probably evolved under similar conditions that occurred naturally or
679 have remarkable plasticity (Wiens et al. 2010, Heino et al. 2013). Finally, our model-based
680 approach can reveal compositional changes. Although the probabilistic adaptation of Jaccard's
681 similarity calculation suggests that predicting individual genera is more challenging than
682 aggregated genus richness, this analysis provided additional information that is generally not
683 easily available from other approaches (Hawkins 2006, Van Sickle 2008) and enhanced our
684 ability to identify sites that are potentially affected by human disturbance.

685 Selecting the appropriate taxonomic level that specimens should be identified to is a
686 critical decision in biological assessment (Chessman et al. 2007). We focused exclusively on
687 taxa that were identified to the genus level and excluded those that could not be unambiguously
688 identified. This possibly increases false absences rates in our study, but NRSA identifies most
689 organisms to their genus such that these instances are relatively rare. Further, common taxonomy

690 alleviates some ambiguity associated with aggregating unresolved taxa into operational
691 taxonomic units (Yuan et al. 2008) and improves transferability of our genus-environment
692 relationships to other studies. Fixed-count subsampling, performed as part of the standardized
693 NRSA protocol, could also increase false absences in our study. Although, explicitly modeling
694 taxon-specific detection probabilities from replicate subsamples may be an interesting avenue for
695 future research (Doser et al. 2023, Doser et al. 2024), these data are presently unavailable.
696 Finally, focusing on genera could also reduce interspecific variation, but species-level taxonomy
697 was not available for this dataset. In general, finer taxonomic resolution would be substantially
698 more expensive and perhaps only yield marginal benefits (Chessman et al. 2007). Nonetheless,
699 metabarcoding approaches may have shown promise for bioassessments (Smucker et al. 2024)
700 and could avoid some of the limitations associated with selecting the appropriate taxonomic
701 resolution.

702 Although the model-based approaches may have some advantages over reference site-
703 based approaches, they require further investigation before they can be fully assimilated into
704 biomonitoring programs. Nevertheless, the National Rivers and Streams Assessment reports
705 biological condition estimates at for the 2018-19 survey using a multi-metric index (USEPA
706 2024b) and it is worthwhile to compare our results to those. The results from NRSA are available
707 at <https://riverstreamassessment.epa.gov/dashboard>. Based on a ranked comparison, the largest
708 disagreements pertained to the percentage of stream miles in poor condition in the Plains
709 ecoregions. Specifically, our model-based approach showed more streams to be altered in the
710 Northern Plains, Southern Plains and Temperate Plains. Indeed, the US Great Plains (i.e.
711 Temperate Plains, Northern Plains and Southern Plains) have undergone extensive conversion
712 from grasslands to agriculture such that there may be few sites that are undisturbed (Samson and

713 Knopf 1994, Dodds et al. 2004, Olimb and Robinson 2019). Because of the large extent of
714 anthropogenic activities in these regions, it is likely that the reference sites used to assess
715 biological condition are potentially lower quality (Herlihy et al. 2008). Conversely, NRSA
716 ranked the Coastal Plains as having a larger percentage of streams in poor condition than our
717 model-based approach. Unfortunately, because of potential extrapolation associated with our
718 model-based assessment, the Coastal Plains also had a relatively large number of streams that
719 could not be assessed. The rankings for the remaining ecoregions tended be similar and both
720 methods agreed that streams in the Western Mountains were the least-disturbed among the
721 ecoregions. Indeed, about 75% of the land area of this ecoregion is in federal ownership, which
722 could convey some protection to streams and rivers (Jenkins et al. 2015). Furthermore, Western
723 Mountains have a relatively shorter history of anthropogenic disturbance than other ecoregions in
724 the United States.

725

726 *Macroinvertebrate response to anthropogenic disturbance*

727 Beyond applications for bioassessment, our model-based approach contributes to a
728 growing literature devoted to understanding of recent trends in macroinvertebrates (Crossley et
729 al. 2020, Jähnig et al. 2021, Gebert et al. 2022, Spake et al. 2022, Rumschlag et al. 2023). In our
730 study, hindcasting estimates tended to show increases in genus richness, suggesting that in many
731 ecoregions fewer genera persist when physiochemical conditions are disturbed. Other studies
732 have reported either modest increases in macroinvertebrate richness (Gebert et al. 2022,
733 Rumschlag et al. 2023) or no clear evidence for a widespread decline of insects (Crossley et al.
734 2020). Because few datasets span more than several decades, an advantage of our space-for-time
735 approach is that it could reflect a period before major anthropogenic disturbances (Blüthgen et al.

736 2022). In addition, the survey design and consistent methodology implemented by NRSA
737 reduces the potential for confounding spatial and temporal dimensions that could be present in
738 datasets that were compiled from multiple sources (Jähnig et al. 2021, Blüthgen et al. 2022,
739 Boyd et al. 2023). Nonetheless, although our results suggest that sites typically support fewer
740 macroinvertebrate genera under present-day conditions, we did detect increased in genus
741 richness at several sites in the Northern Plains and Western Mountains which suggests that
742 macroinvertebrate response to disturbance could be context dependent (Powell et al. 2023).

743 We compared genus-specific occurrence probabilities under present-day and hindcasted
744 conditions to identify which genera contribute to assemblage level patterns. Insects in most
745 ecoregions tended to decrease with disturbance while non-insects tended to increase. This
746 suggests that insect genera may, in general, be less tolerant to anthropogenic disturbance (Jähnig
747 et al. 2021) and aligns with other studies that have documented declines in insect richness
748 (Rumschlag et al. 2023). Among insects, Ephemeroptera, Plecoptera and Trichoptera (EPT) are
749 often used as indicators of anthropogenic disturbance (Stoddard et al. 2008) and may have
750 already lost a considerable proportion of species (Sánchez-Bayo and Wyckhuys 2019). This is
751 especially concerning because we found that EPT occurrence probabilities could continue to
752 decrease in many regions of the US. Among non-insects, non-arthropods included members of
753 Annelida, Nemertea, and Platyhelminthes tended to increase at a relatively large proportion of
754 sites which is consistent with high pollution tolerance values typically assigned to these taxa
755 (Carlisle et al. 2007, Griffith 2023).

756 Because our approach was based on quantifying genus-environment relationships along
757 environmental gradients that are typically altered by human activity, we could also compare the
758 relative effects of changing specific categories of physiochemical variables on macroinvertebrate

759 assemblages. For example, higher genus richness under present-day conditions in the Northern
760 Plains compared to hindcasted habitat variables suggests that disturbance may have increased
761 genera richness. Interestingly, cattle in pasture and rangelands were among the most common
762 types of disturbance measured by RPDI for this region (USEPA 2023) and could potentially
763 increase genera richness by suppressing riparian forest cover and, in turn, elevating primary
764 production (Mittelbach et al. 2001, Tonkin et al. 2013). Conversely, in the Northern
765 Appalachians and Coastal Plains, where the presence of trash, landfills or buildings were the
766 most common factors for RPDI, we found that genus richness was presently lower which is
767 consistent with the negative effects of urbanization on stream macroinvertebrate assemblages
768 (Morse et al. 2003). We also found different effects with respect to changes in precipitation
769 whereby genus richness increased in the Northern Plains and decreased in Southern
770 Appalachians. Collectively, these patterns reveal which environmental variables may be most
771 important in structuring microinvertebrate assemblages and in what contexts – an interpretation
772 that would not be possible with traditional reference site-based approaches.

773

774 *Conclusions*

775 We used a model-based approach to assess the potential effects of anthropogenic
776 disturbance on physiochemical gradients and benthic macroinvertebrate assemblages. Our
777 approach combines genus-environment relationships with estimates of a number of important
778 dimensions in physiochemical condition after removing anthropogenic disturbance. The number
779 of sites in minimally disturbed condition are progressively diminishing, so methods that
780 circumvent the need for reference sites for biological assessments of streams and rivers are
781 crucial to understanding the extent of anthropogenic impacts. Importantly, our framework could

782 provide an avenue to conduct biological assessment without depending on least disturbed
783 reference sites.

784

785 Acknowledgements

786 We thank S. Paulsen, L. Riato, E. Fergus, K. Blocksom, and D. Peck for valuable
787 discussions throughout the development of this project. D. Benkendorf, S. Rumschlag, D.
788 Carlisle and an anonymous reviewer provided comments on earlier versions of this manuscript.
789 The information in this document has been funded entirely by the U.S. Environmental Protection
790 Agency, in part through appointments to the Internship/Research Participation Program at the
791 Office of Research and Development, U.S. Environmental Protection Agency, administered by
792 the Oak Ridge Institute for Science and Education through an interagency agreement. The views
793 expressed in this article are those of the authors and do not necessarily represent the views or
794 policies of the U.S. Environmental Protection Agency. Any mention of trade names, products or
795 services does not imply an endorsement by the U.S. Government or the U.S. Environmental
796 Protection Agency. The EPA does not endorse and commercial products, services, or enterprise
797

798 References:

799 Abrego, N., and O. Ovaskainen. 2023. Evaluating the predictive performance of presence–
800 absence models: Why can the same model appear excellent or poor? *Ecology and
801 Evolution* **13**:e10784.

802 Blüthgen, N., M. Staab, R. Achury, and W. W. Weisser. 2022. Unravelling insect declines: can
803 space replace time? *Biology Letters* **18**:20210666.

- 804 Bouleau, G., and D. Pont. 2015. Did you say reference conditions? Ecological and socio-
805 economic perspectives on the European Water Framework Directive. *Environmental
806 Science & Policy* **47**:32-41.
- 807 Boyd, R. J., G. D. Powney, and O. L. Pescott. 2023. We need to talk about nonprobability
808 samples. *Trends in Ecology & Evolution* **38**:521-531.
- 809 Breiman, L. 2001. Random Forests. *Machine Learning* **45**:5-32.
- 810 Buss, D. F., D. M. Carlisle, T.-S. Chon, J. Culp, J. S. Harding, H. E. Keizer-Vlek, W. A.
811 Robinson, S. Strachan, C. Thirion, and R. M. Hughes. 2014. Stream biomonitoring using
812 macroinvertebrates around the globe: a comparison of large-scale programs.
813 *Environmental Monitoring and Assessment* **187**:4132.
- 814 Calabrese, J. M., G. Certain, C. Kraan, and C. F. Dormann. 2014. Stacking species distribution
815 models and adjusting bias by linking them to macroecological models. *Global Ecology
816 and Biogeography* **23**:99-112.
- 817 Carlisle, D. M., M. R. Meador, S. R. Moulton, and P. M. Ruhl. 2007. Estimation and application
818 of indicator values for common macroinvertebrate genera and families of the United
819 States. *Ecological Indicators* **7**:22-33.
- 820 Chessman, B., S. Williams, and C. Besley. 2007. Bioassessment of streams with
821 macroinvertebrates: effect of sampled habitat and taxonomic resolution. *Journal of the
822 North American Benthological Society* **26**:546-565.
- 823 Chessman, B. C., and M. J. Royal. 2004. Bioassessment without reference sites: use of
824 environmental filters to predict natural assemblages of river macroinvertebrates. *Journal
825 of the North American Benthological Society* **23**:599-615.

- 826 Crossley, M. S., A. R. Meier, E. M. Baldwin, L. L. Berry, L. C. Crenshaw, G. L. Hartman, D.
- 827 Lagos-Kutz, D. H. Nichols, K. Patel, S. Varriano, W. E. Snyder, and M. D. Moran. 2020.
- 828 No net insect abundance and diversity declines across US Long Term Ecological
- 829 Research sites. *Nature Ecology & Evolution* **4**:1368-1376.
- 830 Deflem, I. S., E. Bennetsen, Ø. H. Opedal, F. C. F. Calboli, O. Ovaskainen, G. Van Thuyne, F.
- 831 A. M. Volckaert, and J. A. M. Raeymaekers. 2021. Predicting fish community responses
- 832 to environmental policy targets. *Biodiversity and Conservation* **30**:1457-1478.
- 833 Dodds, W. K., K. Gido, M. R. Whiles, K. M. Fritz, and W. J. Matthews. 2004. Life on the Edge:
- 834 The Ecology of Great Plains Prairie Streams. *BioScience* **54**:205-216.
- 835 Dodds, W. K., and R. M. Oakes. 2004. A technique for establishing reference nutrient
- 836 concentrations across watersheds affected by humans. *Limnology and Oceanography:*
- 837 *Methods* **2**:333-341.
- 838 Doser, J. W., A. O. Finley, and S. Banerjee. 2023. Joint species distribution models with
- 839 imperfect detection for high-dimensional spatial data. *Ecology* **104**:e4137.
- 840 Doser, J. W., A. O. Finley, M. Kéry, and E. F. Zipkin. 2024. spAbundance: An R package for
- 841 single-species and multi-species spatially explicit abundance models. *Methods in*
- 842 *Ecology and Evolution* **n/a**.
- 843 Dudgeon, D., A. H. Arthington, M. O. Gessner, Z. Kawabata, D. J. Knowler, C. Lévéque, R. J.
- 844 Naiman, A. H. Prieur-Richard, D. Soto, M. L. Stiassny, and C. A. Sullivan. 2006.
- 845 Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol*
- 846 *Rev Camb Philos Soc* **81**:163-182.
- 847 Dumelle, M., T. Kincaid, A. R. Olsen, and M. Weber. 2023. spsurvey: Spatial Sampling Design
- 848 and Analysis in R. *Journal of Statistical Software* **105**:1 - 29.

- 849 Elias, C. L., A. R. Calapez, S. F. P. Almeida, B. Chessman, N. Simões, and M. J. Feio. 2016.
- 850 Predicting reference conditions for river bioassessment by incorporating boosted trees in
- 851 the environmental filters method. *Ecological Indicators* **69**:239-251.
- 852 Elith, J., and J. R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and
- 853 Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and*
- 854 *Systematics* **40**:677-697.
- 855 Fergus, C. E., J. R. Brooks, P. R. Kaufmann, A. T. Herlihy, R. A. Hill, R. M. Mitchell, and P.
- 856 Ringold. 2023. Disentangling natural and anthropogenic effects on benthic
- 857 macroinvertebrate assemblages in western US streams. *Ecosphere* **14**:e4688.
- 858 Franklin, J. 2010. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge
- 859 University Press, Cambridge.
- 860 Gebert, F., M. K. Obrist, R. Siber, F. Altermatt, K. Bollmann, and N. Schuwirth. 2022. Recent
- 861 trends in stream macroinvertebrates: warm-adapted and pesticide-tolerant taxa increase in
- 862 richness. *Biology Letters* **18**:20210513.
- 863 Griffith, M. B. 2023. Development of national stressor-specific genus sensitivity values. *Science*
- 864 of The Total Environment **895**:165121.
- 865 Guisan, A., W. Thuiller, and N. E. Zimmermann. 2017. *Habitat Suitability and Distribution*
- 866 *Models: With Applications in R*. Cambridge University Press, Cambridge.
- 867 Hawkins, C. P. 2006. QUANTIFYING BIOLOGICAL INTEGRITY BY TAXONOMIC
- 868 COMPLETENESS: ITS UTILITY IN REGIONAL AND GLOBAL ASSESSMENTS.
- 869 *Ecological Applications* **16**:1277-1294.

- 870 Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000. DEVELOPMENT AND
871 EVALUATION OF PREDICTIVE MODELS FOR MEASURING THE BIOLOGICAL
872 INTEGRITY OF STREAMS. *Ecological Applications* **10**:1456-1477.
- 873 Hawkins, C. P., J. R. Olson, and R. A. Hill. 2010. The reference condition: predicting
874 benchmarks for ecological and water-quality assessments. *Journal of the North American
875 Benthological Society* **29**:312-343.
- 876 Heino, J., D. Schmera, and T. Erős. 2013. A macroecological perspective of trait patterns in
877 stream communities. *Freshwater Biology* **58**:1539-1555.
- 878 Herlihy, A. T., P. R. Kaufmann, and M. E. Mitch. 1991. Stream chemistry in the eastern United
879 States: 2. Current sources of acidity in acidic and low acid-neutralizing capacity streams.
880 *Water Resources Research* **27**:629-642.
- 881 Herlihy, A. T., S. G. Paulsen, J. V. Sickie, J. L. Stoddard, C. P. Hawkins, and L. L. Yuan. 2008.
882 Striving for consistency in a national assessment: the challenges of applying a reference-
883 condition approach at a continental scale. *Journal of the North American Benthological
884 Society* **27**:860-877.
- 885 Herlihy, A. T., and J. C. Sifneos. 2008. Developing nutrient criteria and classification schemes
886 for wadeable streams in the conterminous US. *Journal of the North American
887 Benthological Society* **27**:932-948.
- 888 Herlihy, A. T., J. L. Stoddard, and C. B. Johnson. 1998. The Relationship Between Stream
889 Chemistry and Watershed Land Cover Data in the Mid-Atlantic Region, U.S. Water, Air,
890 and Soil Pollution **105**:377-386.
- 891 Hill, R. A., M. H. Weber, S. G. Leibowitz, A. R. Olsen, and D. J. Thornbrugh. 2016. The
892 Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the

- 893 Conterminous United States. JAWRA Journal of the American Water Resources
894 Association **52**:120-128.
- 895 Hughes, R. M., and D. V. Peck. 2008. Acquiring data for large aquatic resource surveys: the art
896 of compromise among science, logistics, and reality. Journal of the North American
897 Benthological Society **27**:837-859.
- 898 Jähnig, S. C., V. Baranov, F. Altermatt, P. Cranston, M. Friedrichs-Manthey, J. Geist, F. He, J.
899 Heino, D. Hering, F. Hölker, J. Jourdan, G. Kalinkat, J. Kiesel, F. Leese, A. Maasri, M.
900 T. Monaghan, R. B. Schäfer, K. Tockner, J. D. Tonkin, and S. Domisch. 2021. Revisiting
901 global trends in freshwater insect biodiversity. WIREs Water **8**:e1506.
- 902 Jenkins, C. N., K. S. Van Houtan, S. L. Pimm, and J. O. Sexton. 2015. US protected lands
903 mismatch biodiversity priorities. Proceedings of the National Academy of Sciences
904 **112**:5081-5086.
- 905 Johnson, A. A., M. Q. Ott, and M. Dogucu. 2022. Bayes Rules!: An Introduction to Applied
906 Bayesian Modeling Chapman and Hall/CRC.
- 907 Kaufmann, P. R., R. M. Hughes, S. G. Paulsen, D. V. Peck, C. W. Seeliger, T. Kincaid, and R.
908 M. Mitchell. 2022a. Physical habitat in conterminous US streams and Rivers, part 2: A
909 quantitative assessment of habitat condition. Ecological Indicators **141**:109047.
- 910 Kaufmann, P. R., R. M. Hughes, S. G. Paulsen, D. V. Peck, C. W. Seeliger, M. H. Weber, and R.
911 M. Mitchell. 2022b. Physical habitat in conterminous US streams and rivers, Part 1:
912 Geoclimatic controls and anthropogenic alteration. Ecological Indicators **141**:109046.
- 913 Kaufmann, P. R., P. Levine, E.G. Robison, C. Seeliger, and D.V. Peck. 1999. Quantifying
914 Physical Habitat in Wadeable Streams. EPA/620/R-99/003, U.S. Environmental
915 Protection Agency, Washington, D.C.

- 916 Kaushal, S. S., G. E. Likens, P. M. Mayer, R. R. Shatkay, S. A. Shelton, S. B. Grant, R. M. Utz,
917 A. M. Yaculak, C. M. Maas, J. E. Reimer, S. V. Bhide, J. T. Malin, and M. A. Rippy.
918 2023. The anthropogenic salt cycle. *Nature Reviews Earth & Environment* **4**:770-784.
- 919 Kilgour, B., and L. Stanfield. 2005. A new approach for defining reference conditions in streams.
920 AGU Spring Meeting Abstracts.
- 921 Kopp, D. A., J. L. Stoddard, R. A. Hill, J. M. Doyle, P. R. Kaufmann, A. T. Herlihy, and S. G.
922 Paulsen. 2023. Joint species distribution models reveal taxon-specific sensitivities to
923 potential anthropogenic alteration. *Freshwater Science* **42**:268-284.
- 924 Lin, J., J. E. Compton, R. A. Hill, A. T. Herlihy, R. D. Sabo, J. R. Brooks, M. Weber, B. Pickard,
925 S. G. Paulsen, and J. L. Stoddard. 2021. Context is Everything: Interacting Inputs and
926 Landscape Characteristics Control Stream Nitrogen. *Environmental Science &*
927 *Technology* **55**:7890-7899.
- 928 Linke, S., R. H. Norris, D. P. Faith, and D. Stockwell. 2005. ANNA: A new prediction method
929 for bioassessment programs. *Freshwater Biology* **50**:147-158.
- 930 McNellie, M. J., I. Oliver, J. Dorrough, S. Ferrier, G. Newell, and P. Gibbons. 2020. Reference
931 state and benchmark concepts for better biodiversity conservation in contemporary
932 ecosystems. *Global Change Biology* **26**:6702-6714.
- 933 Meinshausen, N. 2017. quantregForest: Quantile Regression Forests. R package version 1.3-7.
- 934 Meyer, H., and E. Pebesma. 2021. Predicting into unknown space? Estimating the area of
935 applicability of spatial prediction models. *Methods in Ecology and Evolution* **12**:1620-
936 1633.

- 937 Mitchell, R. M., A. T. Herlihy, and R. M. Hughes. 2025. Striving for consistency in a national
938 lake assessment: Defining reference status and littoral macroinvertebrate condition in
939 lakes across the conterminous United States. *Ecological Indicators* **170**:112992.
- 940 Mittelbach, G. G., C. F. Steiner, S. M. Scheiner, K. L. Gross, H. L. Reynolds, R. B. Waide, M.
941 R. Willig, S. I. Dodson, and L. Gough. 2001. WHAT IS THE OBSERVED
942 RELATIONSHIP BETWEEN SPECIES RICHNESS AND PRODUCTIVITY? *Ecology*
943 **82**:2381-2396.
- 944 Morse, C. C., A. D. Huryn, and C. Cronan. 2003. Impervious Surface Area as a Predictor of the
945 Effects of Urbanization on Stream Insect Communities in Maine, U.S.A. *Environmental
946 Monitoring and Assessment* **89**:95-127.
- 947 O'brien, R. M. 2007. A Caution Regarding Rules of Thumb for Variance Inflation Factors.
948 *Quality & Quantity* **41**:673-690.
- 949 Olimb, S. K., and B. Robinson. 2019. Grass to grain: Probabilistic modeling of agricultural
950 conversion in the North American Great Plains. *Ecological Indicators* **102**:237-245.
- 951 Olsen, A. R., and D. V. Peck. 2008. Survey design and extent estimates for the Wadeable
952 Streams Assessment. *Journal of the North American Benthological Society* **27**:822-836.
- 953 Olson, J. R., and S. M. Cormier. 2019. Modeling Spatial and Temporal Variation in Natural
954 Background Specific Conductivity. *Environmental Science & Technology* **53**:4316-4325.
- 955 Olson, J. R., and C. Hawkins, P. . 2013. Developing site-specific nutrient criteria from empirical
956 models. *Freshwater Science* **32**:719-740.
- 957 Olson, J. R., and C. P. Hawkins. 2012. Predicting natural base-flow stream water chemistry in
958 the western United States. *Water Resources Research* **48**.

- 959 Outhwaite, C. L., P. McCann, and T. Newbold. 2022. Agriculture and climate change are
960 reshaping insect biodiversity worldwide. *Nature* **605**:97-102.
- 961 Ovaskainen, O., and N. Abrego. 2020. Joint Species Distribution Modelling: With Applications
962 in R. Cambridge University Press, Cambridge.
- 963 Ovaskainen, O., D. B. Roy, R. Fox, and B. J. Anderson. 2016. Uncovering hidden spatial
964 structure in species communities with spatially explicit joint species distribution models.
965 *Methods in Ecology and Evolution* **7**:428-436.
- 966 Ovaskainen, O., G. Tikhonov, A. Norberg, F. Guillaume Blanchet, L. Duan, D. Dunson, T.
967 Roslin, and N. Abrego. 2017. How to make more out of community data? A conceptual
968 framework and its implementation as models and software. *Ecology Letters* **20**:561-576.
- 969 Paulsen, S. G., A. Mayio, D. V. Peck, J. L. Stoddard, E. Tarquinio, S. M. Holdsworth, J. V.
970 Sickle, L. L. Yuan, C. P. Hawkins, A. T. Herlihy, P. R. Kaufmann, M. T. Barbour, D. P.
971 Larsen, and A. R. Olsen. 2008. Condition of stream ecosystems in the US: an overview of
972 the first national assessment. *Journal of the North American Benthological Society*
973 **27**:812-821.
- 974 Peters, J., B. D. Baets, N. E. C. Verhoest, R. Samson, S. Degroeve, P. D. Becker, and W.
975 Huybrechts. 2007. Random forests as a tool for ecohydrological distribution modelling.
976 *Ecological Modelling* **207**:304-318.
- 977 Powell, K. E., T. H. Oliver, T. Johns, M. González-Suárez, J. England, and D. B. Roy. 2023.
978 Abundance trends for river macroinvertebrates vary across taxa, trophic group and river
979 typology. *Glob Chang Biol* **29**:1282-1295.

- 980 Prasad, A. M., L. R. Iverson, and A. Liaw. 2006. Newer Classification and Regression Tree
981 Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* **9**:181-
982 199.
- 983 Reynoldson, T. B., R. H. Norris, V. H. Resh, K. E. Day, and D. M. Rosenberg. 1997. The
984 Reference Condition: A Comparison of Multimetric and Multivariate Approaches to
985 Assess Water-Quality Impairment Using Benthic Macroinvertebrates. *Journal of the*
986 *North American Benthological Society* **16**:833-852.
- 987 Rumschlag, S. L., M. B. Mahon, D. K. Jones, W. Battaglin, J. Behrens, E. S. Bernhardt, P.
988 Bradley, E. Brown, F. De Laender, R. Hill, S. Kunz, S. Lee, E. Rosi, R. Schäfer, T. S.
989 Schmidt, M. Simonin, K. Smalling, K. Voss, and J. R. Rohr. 2023. Density declines,
990 richness increases, and composition shifts in stream macroinvertebrates. *Science*
991 *Advances* **9**:eadf4896.
- 992 Sabo, R. D., C. M. Clark, J. Bash, D. Sobota, E. Cooter, J. P. Dobrowolski, B. Z. Houlton, A.
993 Rea, D. Schwede, S. L. Morford, and J. E. Compton. 2019. Decadal Shift in Nitrogen
994 Inputs and Fluxes Across the Contiguous United States: 2002–2012. *Journal of*
995 *Geophysical Research: Biogeosciences* **124**:3104-3124.
- 996 Sabo, R. D., C. M. Clark, D. A. Gibbs, G. S. Metson, M. J. Todd, S. D. LeDuc, D. Greiner, M.
997 M. Fry, R. Polinsky, Q. Yang, H. Tian, and J. E. Compton. 2021. Phosphorus Inventory
998 for the Conterminous United States (2002–2012). *Journal of Geophysical Research:*
999 *Biogeosciences* **126**:e2020JG005684.
- 1000 Sabo, R. D., B. Pickard, J. Lin, B. Washington, C. M. Clark, J. E. Compton, M. Pennino, B.
1001 Bierwagen, S. D. LeDuc, J. N. Carleton, M. Weber, M. Fry, R. Hill, S. Paulsen, A.
1002 Herlihy, and J. L. Stoddard. 2023. Comparing Drivers of Spatial Variability in U.S. Lake

- 1003 and Stream Phosphorus Concentrations. *Journal of Geophysical Research: Biogeosciences* **128**:e2022JG007227.
- 1004
- 1005 Samson, F., and F. Knopf. 1994. Prairie conservation in North America. *BioScience* **44**:418-421.
- 1006 Sánchez-Bayo, F., and K. A. G. Wyckhuys. 2019. Worldwide decline of the entomofauna: A
- 1007 review of its drivers. *Biological Conservation* **232**:8-27.
- 1008 Scherrer, D., H. K. Mod, and A. Guisan. 2020. How to evaluate community predictions without
- 1009 thresholding? *Methods in Ecology and Evolution* **11**:51-63.
- 1010 Smucker, N. J., E. M. Pilgrim, C. T. Nietch, L. Gains-Germain, C. Carpenter, J. A. Darling, L. L.
- 1011 Yuan, R. M. Mitchell, and A. I. Pollard. 2024. Using DNA metabarcoding to characterize
- 1012 national scale diatom-environment relationships and to develop indicators in streams and
- 1013 rivers of the United States. *Science of The Total Environment* **939**:173502.
- 1014 Soranno, P. A., T. Wagner, S. L. Martin, C. McLean, L. N. Novitski, C. D. Provence, and A. R.
- 1015 Rober. 2011. Quantifying regional reference conditions for freshwater ecosystem
- 1016 management: A comparison of approaches and future research needs. *Lake and Reservoir*
- 1017 *Management* **27**:138-148.
- 1018 Spake, R., M. P. Barajas-Barbosa, S. A. Blowes, D. E. Bowler, C. T. Callaghan, M. Garbowski,
- 1019 S. D. Jurburg, R. van Klink, L. Korell, E. Ladouceur, R. Rozzi, D. S. Viana, W.-B. Xu,
- 1020 and J. M. Chase. 2022. Detecting Thresholds of Ecological Change in the Anthropocene.
- 1021 *Annual Review of Environment and Resources* **47**:797-821.
- 1022 Stoddard, J. L., A. T. Herlihy, D. V. Peck, R. M. Hughes, T. R. Whittier, and E. Tarquinio. 2008.
- 1023 A process for creating multimetric indices for large-scale aquatic surveys. *Journal of the*
- 1024 *North American Benthological Society* **27**:878-891.

- 1025 Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. SETTING
1026 EXPECTATIONS FOR THE ECOLOGICAL CONDITION OF STREAMS: THE
1027 CONCEPT OF REFERENCE CONDITION. *Ecological Applications* **16**:1267-1276.
- 1028 Tang, T., R. J. Stevenson, and J. B. Grace. 2020. The importance of natural versus human factors
1029 for ecological conditions of streams and rivers. *Science of The Total Environment*
1030 **704**:135268.
- 1031 Tikhonov, G., Ø. H. Opedal, N. Abrego, A. Lehikoinen, M. M. J. de Jonge, J. Oksanen, and O.
1032 Ovaskainen. 2020. Joint species distribution modelling with the r-package Hmsc.
1033 *Methods in Ecology and Evolution* **11**:442-447.
- 1034 Tonkin, J. D., R. G. Death, and K. J. Collier. 2013. Do productivity and disturbance interact to
1035 modulate macroinvertebrate diversity in streams? *Hydrobiologia* **701**:159-172.
- 1036 USEPA. 2017a. National Rivers and Streams Assessment 2018-19: Field Operations Manual–
1037 Wadeable. EPA-841-B-17-003a. U.S. Environmental Protection Agency, Office of
1038 Water, Washington, DC. .
- 1039 USEPA. 2017b. National Rivers and Streams Assessment 2018-19: Laboratory Operations
1040 Manual. EPA841-B-17-004. U.S. Environmental Protection Agency, Office of Water,
1041 Washington, DC.
- 1042 USEPA. 2023. National Rivers and Streams Assessment 2018-2019 Technical Support
1043 Document. EPA 841-R-22-005. U.S. Environmental Protection Agency, Office of Water
1044 and Office of Research and Development, Washington, D.C. .
- 1045 USEPA. 2024a. Climate change indicators in the Unites States. EPA 430-R-24-003.

- 1046 USEPA. 2024b. National Rivers and Streams Assessment: The Third Collaborative Survey. U.S.
- 1047 Environmental Protection Agency, Office of Water and Office of Research and
- 1048 Development.
- 1049 Van Sickle, J. 2008. An index of compositional dissimilarity between observed and expected
- 1050 assemblages. *Journal of the North American Benthological Society* **27**:227-235.
- 1051 Vörösmarty, C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S.
- 1052 Glidden, S. E. Bunn, C. A. Sullivan, C. R. Liermann, and P. M. Davies. 2010. Global
- 1053 threats to human water security and river biodiversity. *Nature* **467**:555-561.
- 1054 Warton, D. I., F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F.
- 1055 K. C. Hui. 2015. So Many Variables: Joint Modeling in Community Ecology. *Trends in*
- 1056 *Ecology & Evolution* **30**:766-779.
- 1057 Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I.
- 1058 Damschen, T. Jonathan Davies, J.-A. Grytnes, S. P. Harrison, B. A. Hawkins, R. D. Holt,
- 1059 C. M. McCain, and P. R. Stephens. 2010. Niche conservatism as an emerging principle in
- 1060 ecology and conservation biology. *Ecology Letters* **13**:1310-1324.
- 1061 Wilkinson, D. P., N. Golding, G. Guillera-Arroita, R. Tingley, and M. A. McCarthy. 2021.
- 1062 Defining and evaluating predictions of joint species distribution models. *Methods in*
- 1063 *Ecology and Evolution* **12**:394-404.
- 1064 Wood, G. 2008. Thresholds and criteria for evaluating and communicating impact significance in
- 1065 environmental statements: ‘See no evil, hear no evil, speak no evil’? *Environmental*
- 1066 *Impact Assessment Review* **28**:22-38.
- 1067 Yuan, L. L., D. M. Carlisle, R. M. Mitchell, and A. I. Pollard. 2024. Using models of local
- 1068 environmental conditions for biological assessment. *Freshwater Science*:000-000.

- 1069 Yuan, L. L., C. P. Hawkins, and J. V. Sickle. 2008. Effects of regionalization decisions on an
1070 O/E index for the US national assessment. *Journal of the North American Benthological
1071 Society* **27**:892-905.
- 1072 Zak, D., M. Hupfer, A. Cabezas, G. Jurasinski, J. Audet, A. Kleeberg, R. McInnes, S. M.
1073 Kristiansen, R. J. Petersen, H. Liu, and T. Goldhammer. 2021. Sulphate in freshwater
1074 ecosystems: A review of sources, biogeochemical cycles, ecotoxicological effects and
1075 bioremediation. *Earth-Science Reviews* **212**:103446.

1076

For Review Only

1077 Tables

1078 Table 1: Sites, genera, and environmental variables included in the analysis. Revisit sites are locations that were revisited during the
 1079 survey to assess within year variability. Values for the environmental variables are median values for each region. Range is given in
 1080 parentheses. NTL = Total Nitrogen, PTL = Total Phosphorous, CL = Chloride, SO₄ = Sulfate, RPDI = Riparian Disturbance Index,
 1081 SUBD = Substrate Diameter, TPRCP = Total Precipitation and MSAT = Mean Summer Air Temperature. CPL = Coastal Plains,
 1082 NAP = Northern Appalachians, NPL = Northern Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL = Temperate
 1083 Plains, UMW = Upper Midwest; WMT = Western Mountains; XER = Xeric.

1084

Region	Site (#)	Genera (#)	Revisit Sites (#)	NTL (ug/L)	PTL (ug/L)	CL (mg/L)	SO ₄ (mg/L)	RPDI	SUBD (mm)	TPRCP (mm)	MSAT (°C)
CPL	226	71	36	786 (91-7713)	84.81 (5.2-3922.25)	10.06 (0.52-4797.9)	7.83 (0.11-1900.89)	0.5 (0-5.39)	0.35 (0.01-341.65)	1613.83 (388.66-2397.1)	27.46 (23.75-31.68)
NAP	228	127	29	431 (81-6413)	23.76 (3.12-587.2)	16.73 (0.07-668.16)	5.58 (0.06-467.93)	0.65 (0-5.24)	11.66 (0.01-864.9)	1295.41 (888.91-1883.17)	21.49 (17.24-24.72)
NPL	152	76	7	881.5 (71-15675)	85.61 (3.53-9248.31)	12.09 (0.11-1251.58)	486.24 (3.5-4079.78)	1.27 (0-4.49)	0.51 (0.01-560.5)	437.42 (157.84-1055.11)	20.52 (15.62-23.46)
SAP	266	111	32	557.5 (36-18700)	29.3 (3.79-4050)	5.89 (0.36-197.39)	7.46 (0.58-397.69)	0.76 (0-4.56)	16.28 (0.01-5656.85)	1475.78 (889.64-2451.81)	24.76 (19.19-27.73)
SPL	174	59	5	1165 (145-21175)	147.12 (5.21-4351.7)	30.36 (0.43-5220)	88.89 (1.96-3716.6)	1.04 (0-5.88)	0.35 (0.01-5656.85)	616.17 (240.1-1404.61)	25.88 (11.96-32.75)
TPL	223	74	29	1806 (236-16219)	165.08 (11.45-1066.87)	18.9 (1.44-736.62)	34.4 (5.79-1386.77)	0.83 (0-5.47)	0.35 (0.01-5656.85)	1004.22 (376.7-1801.08)	23.33 (18.28-27.09)
UMW	201	104	10	1168 (195-17675)	57.4 (8.16-659)	9.97 (0.01-306.69)	9.28 (0.04-160.4)	0.62 (0-6.43)	0.35 (0.01-812.79)	960.16 (496.44-1718.2)	20.84 (17.05-23.43)
WMT	225	94	19	133 (22-4719)	21.84 (2.71-569.8)	0.92 (0.04-521.64)	2.97 (0.07-1682.05)	0.48 (0-3.95)	38.14 (0.01-1288.61)	669.31 (175.89-3946.33)	18.08 (11.71-29.4)
XER	196	75	16	344.5 (46-8000)	54.62 (4.29-4667.41)	5.35 (0.1-1867.57)	27.8 (0.02-3286.24)	1.16 (0-4.36)	2.02 (0.01-368.11)	287.58 (100.46-1176.56)	22.76 (14.69-36.11)

1086 Table 2: Random Forest model performance metrics for testing and out-of-bag training datasets. NTL = Total Nitrogen, PTL = Total
1087 Phosphorus, CL = Chloride, SO₄ = Sulfate, SUBD = Substrate Diameter. RMSE = Root mean squared error of random forest models
1088 fitted with ln(x + 1) (NTL, PTL, and CL) or ln(x) (SO₄) or Log₁₀(SUBD) transformations.

1089

Variable	R ² _{train}	RMSE _{train}	R ² _{test}	RMSE _{test}
ln(NTL+1) (ug/L)	0.73	0.61	0.72	0.60
ln(PTL+1) (ug/L)	0.60	0.78	0.69	0.70
ln(CL+1) (mg/L)	0.73	0.75	0.68	0.85
ln(SO ₄) (mg/L)	0.77	0.96	0.78	0.97
log ₁₀ (SUBD) (mm)	0.46	0.98	0.51	1.00

1090

1091

1092

1093

1094

1095

1096

1097

1098

1099

1100

1101

1102

1103

1104 Table 3. Regression coefficients between predicted and observed richness and compositional similarity. Occurrence probabilities
 1105 thresholds (Pr) were used to exclude genera with low predicted occurrence probabilities. We considered $R^2 \geq 0.2$, $-1.5 \leq \text{intercept} \leq$
 1106 1.5 , and $0.85 \leq \text{slope} \leq 1.15$ are indicative of adequate model performance. Values in parentheses are the 5th and 95th percentile of
 1107 estimates from the 3000 posterior samples. Compositional similarity was measured using a probabilistic adaptation of Jaccard index
 1108 measured for each site. Parentheses are the 5th and 95th percentile of values for all sites.

	Pr	Genus Richness				Composition	
		Intercept	Slope	R^2	Revisit Sites within Posterior Distribution (%)	Mean Jaccard Similarity	Mean Jaccard Similarity at revisit sites
CPL	0.05	0.37 (-0.37 - 1.02)	1.00 (0.96 - 1.05)	0.85 (0.82 - 0.87)	97	0.52 (0.26 - 0.7)	0.68 (0.24-0.95)
NAP	0.05	-0.45 (-2.02 - 1.02)	1.03 (0.99 - 1.08)	0.74 (0.71 - 0.78)	90	0.52 (0.31 - 0.64)	0.35 (0.17-0.65)
NPL	0.05	-0.07 (-1.21 - 1.04)	1.02 (0.96 - 1.09)	0.76 (0.71 - 0.81)	100	0.58 (0.34 - 0.77)	0.39 (0.12-0.71)
SAP	0.05	0.40 (-0.66 - 1.41)	1.00 (0.97 - 1.04)	0.85 (0.82 - 0.87)	97	0.52 (0.26 - 0.67)	0.56 (0.21-0.87)
SPL	0.10	0.48 (-0.45 - 1.29)	1.02 (0.97 - 1.09)	0.79 (0.75 - 0.83)	100	0.52 (0.25 - 0.72)	0.61 (0.52-0.98)
TPL	0.05	0.06 (-0.91 - 0.94)	1.01 (0.97 - 1.06)	0.82 (0.79 - 0.84)	100	0.55 (0.31 - 0.77)	0.59 (0.18-1)
UMW	0.05	-0.74 (-2.52 - 0.99)	1.04 (0.98 - 1.10)	0.73 (0.66 - 0.78)	100	0.55 (0.34 - 0.69)	0.53 (0.27-0.9)
WMT	0.10	0.24 (-1.24 - 1.59)	1.04 (0.99 - 1.10)	0.74 (0.70 - 0.78)	100	0.60 (0.36 - 0.75)	0.45 (0.16-0.79)
XER	0.05	-0.12 (-1.11 - 0.74)	1.03 (0.98 - 1.08)	0.83 (0.80 - 0.86)	94	0.50 (0.23 - 0.71)	0.48 (0.07-0.83)

1109

1110

1111

1112

1113 Table 4: Mean proportion of sites in each taxonomic group that was identified as an increaser or decreaser. Increasers (I) are genera
 1114 that have a significantly higher probability of occurrence under present day conditions compared to hindcasted conditions; decreasers
 1115 (D) are genera that have a significantly lower probability of occurrence under present-day conditions compared to hindcasted
 1116 conditions. The values in the table are the mean proportion of sites for genera within the major taxonomic group. CPL = Coastal
 1117 Plains, NAP = Northern Appalachians, NPL = Northern Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL =
 1118 Temperate Plains, UMW = Upper Midwest; WMT = Western Mountains; XER = xeric. EPT = Ephemeroptera, Plecoptera and
 1119 Trichoptera.

	CPL		NAP		NPL		SAP		SPL		TPL		UMW		WMT		XER	
	D	I	D	I	D	I	D	I	D	I	D	I	D	I	D	I	D	I
Insects	0.16	0.08	0.08	0.05	0.1	0.17	0.1	0.11	0.09	0.12	0.16	0.11	0.11	0.12	0.03	0.06	0.09	0.05
EPT	0.24	0.01	0.07	0.04	0.17	0.11	0.13	0.08	0.09	0.04	0.23	0.04	0.14	0.03	0.04	0.06	0.1	0.03
CHIRONOMIDAE	0.11	0.09	0.09	0.05	0.04	0.18	0.04	0.12	0.02	0.17	0.12	0.14	0.06	0.16	0.02	0.07	0.06	0.1
Other Insects	0.21	0.07	0.08	0.08	0.07	0.21	0.13	0.12	0.23	0.02	0.18	0.11	0.09	0.04	0.02	0.03	0.11	0.01
Non-Insects	0.14	0.16	0.06	0.18	0.06	0.19	0.11	0.13	0.06	0.13	0.15	0.15	0.08	0.14	0.03	0.08	0.04	0.14
ARTHROPODA	0.21	0.11	0.07	0.16	0.11	0.02	0.16	0.09	0.05	0.06	0.23	0.35	0.1	0.04	0.04	0.03	0.05	0.04
MOLLUSCA	0.08	0.11	0.03	0.28	0.02	0.3	0.1	0.15	0.09	0.06	0.21	0.05	0.14	0.07	0.01	0.06	0.01	0.26
Other Non-Insects	0.15	0.22	--	0.14	0.05	0.13	0.04	0.16	0.02	0.19	0.05	0.15	0.01	0.25	0.04	0.14	0.04	0.14

1121 Figure Captions

1122 Figure 1: Ecoregions and survey locations for the National Rivers and Streams Assessment
1123 2018-2019 survey. CPL = Coastal Plains, NAP = Northern Appalachians, NPL = Northern
1124 Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL = Temperate Plains, UMW =
1125 Upper Midwest; WMT = Western Mountains; XER = Xeric

1126

1127 Figure 2: Conceptual diagram of posterior distributions generated from bayesian inference. The
1128 grey is a posterior distribution of genus richness generated after removing anthropogenic
1129 disturbances (Hindcasted) and the white is a posterior distribution generated from present-day
1130 conditions. The dotted line is the expected value (i.e. mean) of the present day posterior distribution
1131 and dark shading represents the lower 10% or the upper 90% of the HC posterior distribution. At
1132 site A, the PD mean is within the <10% of the HC distribution and indicates that PD richness is
1133 likely lower than HC richness. Alternatively, at site B, PD richness is >90% of the HC distribution
1134 and indicates that PD richness is likely higher. At site C, the two posterior distributions are similar
1135 such that there is likely no difference between PD and HC.

1136

1137 Figure 3: Partial dependence plots showing the effects of anthropogenic variables on in-stream
1138 physiochemical factors. For visualization, each anthropogenic factor was rescaled between 0-1
1139 and labeled low, medium, and high. CoalMineDen = coal mine density, MineDen = gravel mine
1140 density, MSAT = mean summer air temperature, N_dep = atmospheric nitrogen deposition,
1141 N_input = anthropogenic nitrogen inputs, P_input = anthropogenic phosphorous inputs, PctCrop
1142 = percent crop in the watershed, PctCropRP = percent crop in the riparian area, PctNatRP =
1143 percent natural vegetation in riparian area, RdDen = road density, S_dep = atmospheric sulfur
1144 deposition, TPRCP = total precipitation, and W1_HAG = agricultural disturbance adjacent to
1145 stream reach. NTL = Total Nitrogen (ug/L), PTL = Total Phosphorus (ug/L), CL = Chloride
1146 (mg/L), SO4 = Sulfate (mg/L), SUBD = Substrate Diameter Log10(mm).

1147

1148 Figure 4: Observed versus hindcasted values for each environmental gradient. Points are regional
1149 means and vertical and horizontal bars represent the 10th and 90th quantiles of observed values or
1150 hindcasted values within each region, respectively. For RPDI, all hindcast values were < 0.33
1151 and the ecoregions were plotted separately for visualization. The dashed line is the 1:1
1152 relationship for all plots except for RPDI where it represents the 0.33 threshold applied to all
1153 ecoregions. NTL = Total Nitrogen (ug/L), PTL = Total Phosphorus (ug/L), CL = Chloride (mg/L),
1154 SO4 = Sulfate (mg/L), SUBD = Substrate Diameter (mm), RPDI = Riparian Disturbance Index,
1155 MSAT = mean summer air temperature (°C) and TPRCP = total precipitation (mm). CPL =
1156 Coastal Plains, NAP = Northern Appalachians, NPL = Northern Plains, SAP = Southern
1157 Appalachians, SPL = Southern Plains, TPL = Temperate Plains, UMW = Upper Midwest; WMT
1158 = Western Mountains; XER = Xeric. See Appendix S1: Table S4 for observed and hindcasted
1159 mean values and quantiles for each region.

1160

1161 Figure 5: A) Proportion of sites where hindcasted abiotic conditions were >2 standard deviations
1162 from the observed value for Total Nitrogen (NTL), Total Phosphorus (PTL), Chloride (CL),
1163 Sulfate (SO₄), Substrate Diameter (SUBD), mean summer air temperature (MSAT), and Total
1164 Precipitation (TPRCP) and > 0.33 for Riparian Disturbance Index (RPDI). B) The number of
1165 environmental variables affected by human disturbance. Locations where the effects were not
1166 detected are plotted separately. CPL = Coastal Plains, NAP = Northern Appalachians, NPL =
1167 Northern Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL = Temperate
1168 Plains, UMW = Upper Midwest; WMT = Western Mountains; XER = Xeric.
1169

1170 Figure 6: Difference in macroinvertebrate genus richness after removing anthropogenic
1171 disturbance from each category of physiochemical variables (A) and all variables simultaneously
1172 (B). The points in each panel are sites that had a change in genus richness with $>75\%$ support
1173 after removing disturbance.

1174

1175 Figure 7: Percentage of the population of streams where genus richness or composition could
1176 change given a hindcasted physiochemical environment. Shaded bars indicate that the probability
1177 of mean present-day richness differs from hindcast with > 0.90 (Black) or 0.75-0.90 (Gray)
1178 support. Hatched bars indicate that Jaccard similarity was < 0.9 but support for a difference in
1179 richness were < 0.75 . White bars indicate the proportion of sites that had < 0.75 support for
1180 change in richness and > 0.9 compositional similarity. Yellow bars indicate the proportion of sites
1181 that could not be assessed because of insufficient data or potential extrapolation from predicting
1182 hindcast physiochemical conditions. Regions are arranged according to richness. NPL =
1183 Northern Plains, TPL = Temperate Plains, SPL = Southern Plains, XER = Xeric, SAP =
1184 Southern Appalachians, CPL = Coastal Plains, NAP = Northern Appalachians, UMW = Upper
1185 Midwest and WMT = Western Mountains.

1186

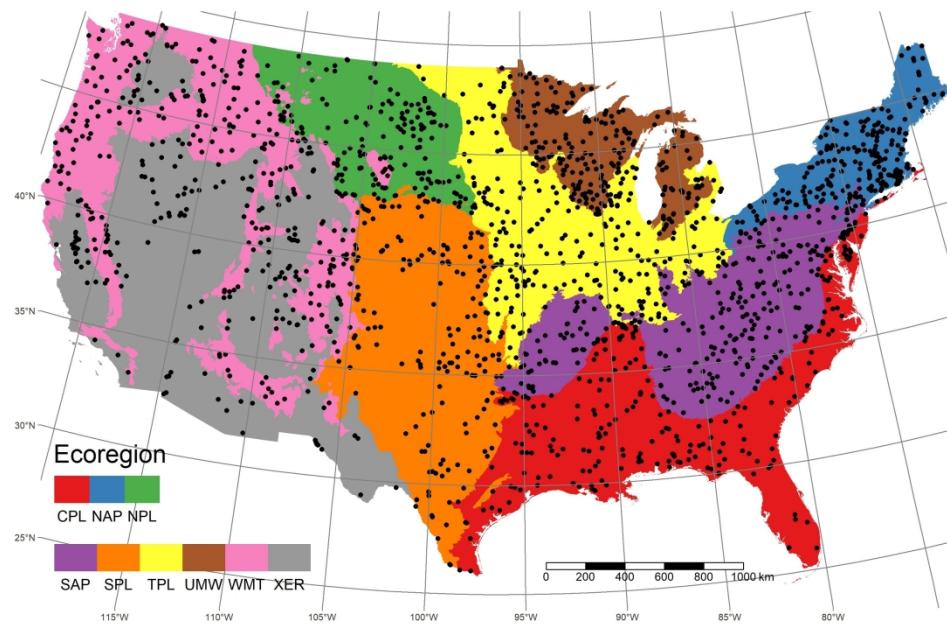


Figure 1: Ecoregions and survey locations for the National Rivers and Streams Assessment 2018-2019 survey. CPL = Coastal Plains, NAP = Northern Appalachians, NPL = Northern Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL = Temperate Plains, UMW = Upper Midwest; WMT = Western Mountains; XER = Xeric

279x177mm (300 x 300 DPI)

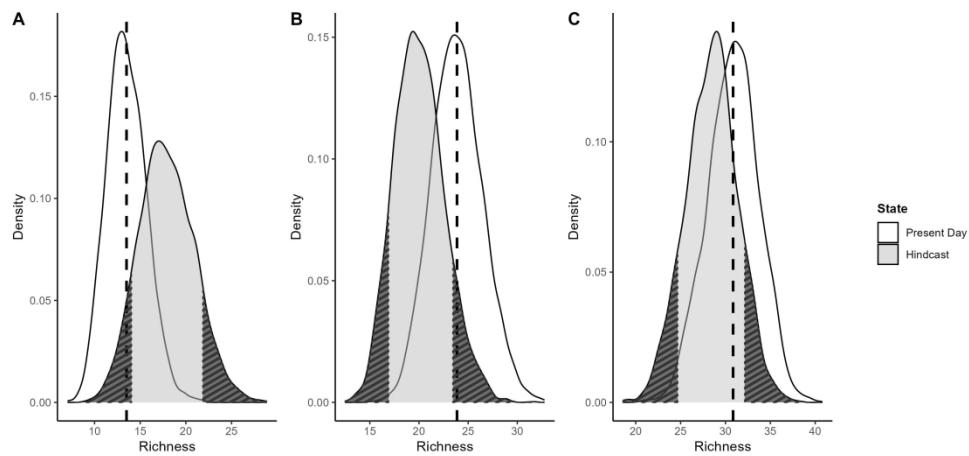


Figure 2: Conceptual diagram of posterior distributions generated from bayesian inference. The grey is a posterior distribution of genus richness generated after removing anthropogenic disturbances (Hindcasted) and the white is a posterior distribution generated from present-day conditions. The dotted line is the expected value (i.e. mean) of the present day posterior distribution and dark shading represents the lower 10% or the upper 90% of the HC posterior distribution. At site A, the PD mean is within the <10% of the HC distribution and indicates that PD richness is likely lower than HC richness. Alternatively, at site B, PD richness is >90% of the HC distribution and indicates that PD richness is likely higher. At site C, the two posterior distributions are similar such that there is likely no difference between PD and HC.

1164x529mm (72 x 72 DPI)

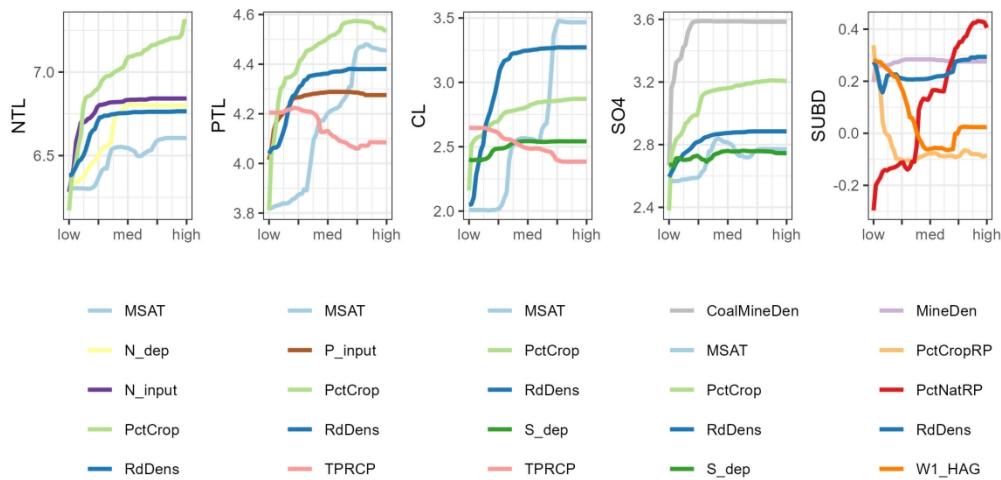


Figure 3: Partial dependence plots showing the effects of anthropogenic variables on in-stream physiochemical factors. For visualization, each anthropogenic factor was rescaled between 0-1 and labeled low, medium, and high. CoalMineDen = coal mine density, MineDen = gravel mine density, MSAT = mean summer air temperature, N_dep = atmospheric nitrogen deposition, N_input = anthropogenic nitrogen inputs, P_input = anthropogenic phosphorous inputs, PctCrop = percent crop in the watershed, PctCropRP = percent crop in the riparian area, PctNatRP = percent natural vegetation in riparian area, RdDen = road density, S_dep = atmospheric sulfur deposition, TPRCP = total precipitation, and W1_HAG = agricultural disturbance adjacent to stream reach. NTL = Total Nitrogen (ug/L), PTL = Total Phosphorus (ug/L), CL = Chloride (mg/L), SO4 = Sulfate (mg/L), SUBD = Substrate Diameter Log10(mm).

846x423mm (72 x 72 DPI)

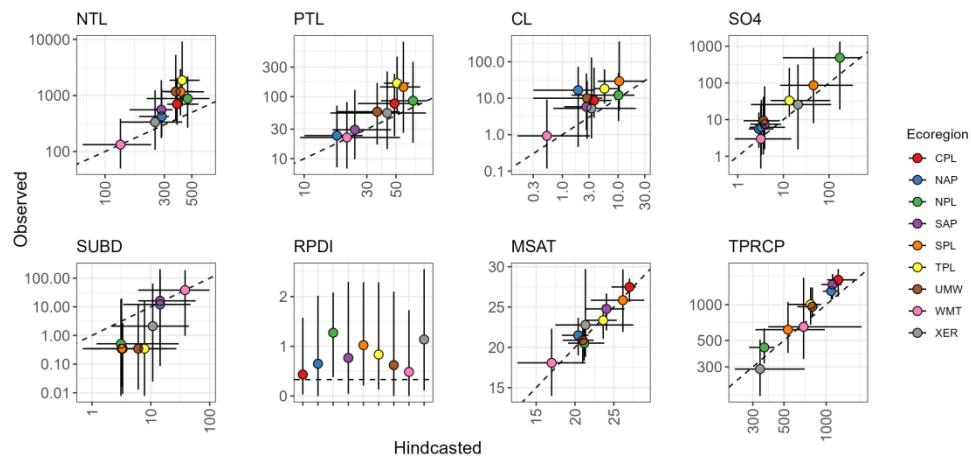


Figure 4: Observed versus hindcasted values for each environmental gradient. Points are regional means and vertical and horizontal bars represent the 10th and 90th quantiles of observed values or hindcasted values within each region, respectively. For RPDI, all hindcast values were < 0.33 and the ecoregions were plotted separately for visualization. The dashed line is the 1:1 relationship for all plots except for RPDI where it represents the 0.33 threshold applied to all ecoregions. NTL = Total Nitrogen (ug/L), PTL = Total Phosphorus (ug/L), CL = Chloride (mg/L), SO4 = Sulfate (mg/L), SUBD = Substrate Diameter (mm), RPDI = Riparian Disturbance Index, MSAT = mean summer air temperature (°C) and TPRCP = total precipitation (mm). CPL = Coastal Plains, NAP = Northern Appalachians, NPL = Northern Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL = Temperate Plains, UMW = Upper Midwest; WMT = Western Mountains; XER = Xeric. See Appendix S1: Table S4 for observed and hindcasted mean values and quantiles for each region.

1164x529mm (72 x 72 DPI)

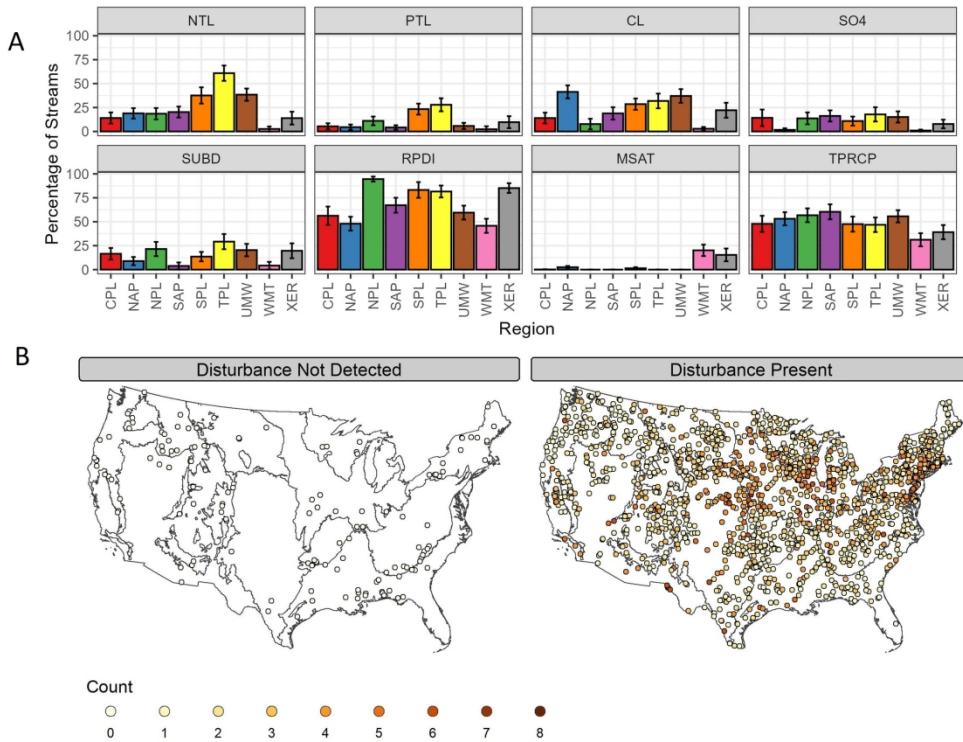


Figure 5: A) Proportion of sites where hindcasted abiotic conditions were >2 standard deviations from the observed value for Total Nitrogen (NTL), Total Phosphorus (PTL), Chloride (CL), Sulfate (SO₄), Substrate Diameter (SUBD), mean summer air temperature (MSAT), and Total Precipitation (TPRCP) and > 0.33 for Riparian Disturbance Index (RPDI). B) The number of environmental variables affected by human disturbance. Locations where the effects were not detected are plotted separately. CPL = Coastal Plains, NAP = Northern Appalachians, NPL = Northern Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL = Temperate Plains, UMW = Upper Midwest; WMT = Western Mountains; XER = Xeric.

218x163mm (300 x 300 DPI)

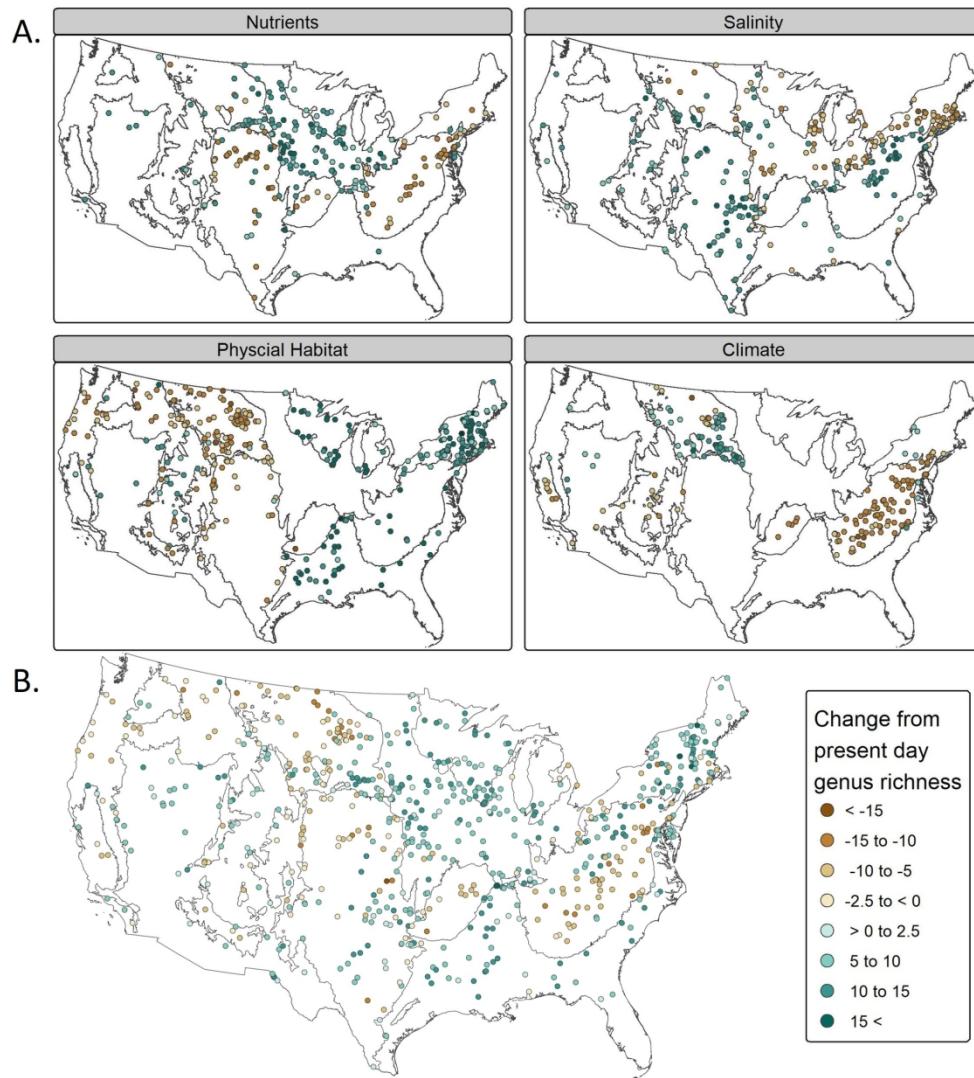


Figure 6: Difference in macroinvertebrate genus richness after removing anthropogenic disturbance from each category of physiochemical variables (A) and all variables simultaneously (B). The points in each panel are sites that had a change in genus richness with >75% support after removing disturbance.

175x191mm (300 x 300 DPI)

Support for benthic macroinvertebrate assemblage change

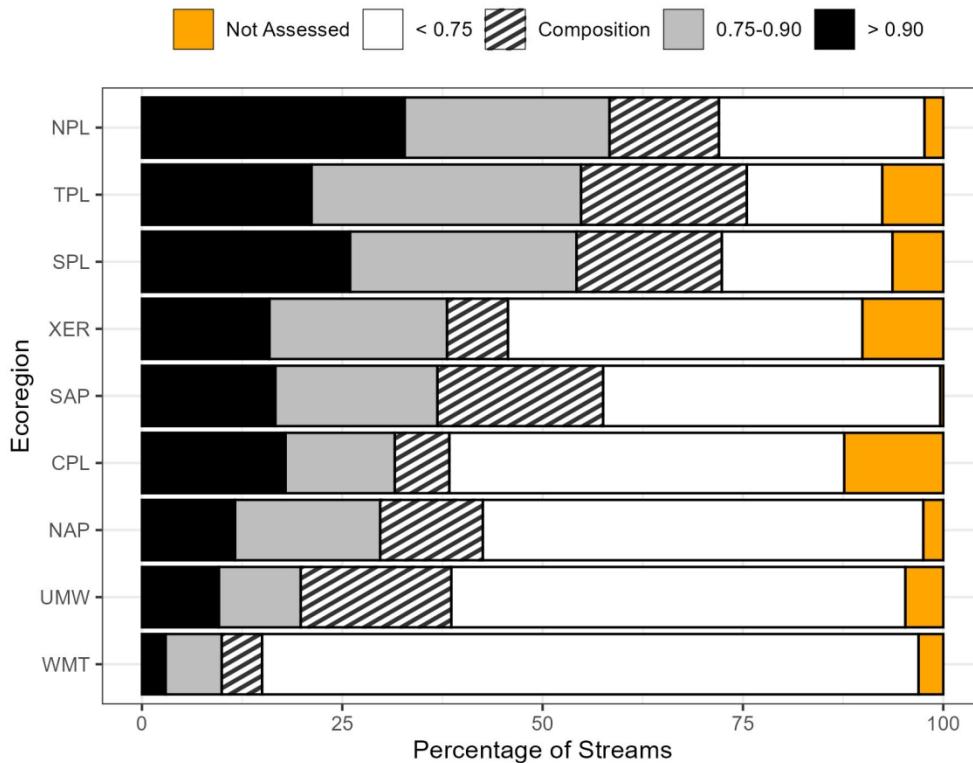


Figure 7: Percentage of the population of streams where genus richness or composition could change given a hindcasted physiochemical environment. Shaded bars indicate that the probability of mean present-day richness differs from hindcast with > 0.90 (Black) or $0.75-0.90$ (Gray) support. Hatched bars indicate that Jaccard similarity was < 0.9 but support for a difference in richness were < 0.75 . White bars indicate the proportion of sites that had < 0.75 support for change in richness and > 0.9 compositional similarity. Yellow bars indicate the proportion of sites that could not be assessed because of insufficient data or potential extrapolation from predicting hindcast physiochemical conditions. Regions are arranged according to richness. NPL = Northern Plains, TPL = Temperate Plains, SPL = Southern Plains, XER = Xeric, SAP = Southern Appalachians, CPL = Coastal Plains, NAP = Northern Appalachians, UMW = Upper Midwest and WMT = Western Mountains.

635x529mm (72 x 72 DPI)

1
2 Supplementary Information
3 A model-based assessment of anthropogenic disturbance on lotic macroinvertebrate assemblages

4
5 Darin A. Kopp^{1*}, John L. Stoddard¹, Philip R. Kaufmann^{1,2}, Alan T. Herlihy², Ryan A. Hill¹,
6 Meredith M. Brehob³, Robert D Sabo⁴

7
8 ¹US EPA, Office of Research and Development, Corvallis, Oregon, USA
9 ²Department of Fisheries, Wildlife, & Conservation Sciences, Oregon State University,
10 Corvallis, Oregon, USA
11 ³Oak Ridge Institute for Science and Education Participant c/o US EPA, ORD, Washington,
12 D.C., USA
13 ⁴US EPA, Office of Research and Development, Center for Public Health and Environmental
14 Assessment, Washington, DC, USA

15
16 *Correspondence: kopp.darin@epa.gov
17
18 Keywords: Joint Species Distribution Modeling, Richness, National Rivers and Streams
19 Assessment, Benthic Macroinvertebrate, Reference Sites, National Aquatic Resource Surveys.

20
21
22
23
24

25 Table S1: Geoclimatic and anthropogenic variables used in random forest models. Variables
26 were selected based on their hypothesized relationship with total nitrogen, total phosphorus,
27 chloride, sulfate and substrate diameter. StreamCat data is publicly accessible at
28 <https://www.epa.gov/national-aquatic-resource-surveys/streamcat-dataset> (Hill et al. 2016) and
29 National Atmospheric Deposition Program Data is publicly accessible at
30 <https://nadp.slh.wisc.edu/maps-data/ntn-gradient-maps/>

For Review Only

Variable	Description	Source
BFIWs	The component of streamflow that can be attributed to ground-water discharge	StreamCat, Hill et al. 2016
ClayWs	Mean % clay content of soils in watershed	StreamCat
ElevWs	Mean elevation of watershed (m)	StreamCat
KffactWs	Mean of STATSGO Kffactor	StreamCat
DepCIWs	Mean atmospheric chloride deposition 2018	National Atmospheric Deposition Program
DepPWs	Atmospheric phosphorous deposition 2013	Sabo et al 2023
NWs	Rock derived nitrogen in watershed	StreamCat
P2O5Ws	Mean % of lithological phosphorous oxide (P2O5) content in surface or near surface geology	StreamCat
PctAlluvCoastWs	% of watershed area classified with lithology type as alluvium and fine-textured coastal zone sediment	StreamCat
PctHbWetWs	% of watershed area classified as herbaceous wetland land cover	StreamCat
PermWs	Mean permeability (cm/hour) of soils (STATSGO) within watershed	StreamCat
RunoffWs	Mean runoff (mm) within watershed	StreamCat
SandWs	Mean % sand content of soils (STATSGO) within watershed	StreamCat
StreamPower	An index reflecting the amount of energy the water exerts on the sides and bottom of a stream	$((TPRCP - ET)/1000) * (0.032 * WSArea)^{0.5} * \text{slope}$
SWs	Mean % of lithological sulfur (S) content in surface or near surface geology within watershed	StreamCat
WsAreaSqKm	Area of watershed (square km)	StreamCat
CoalMineDensWs	Density of coal mines sites within watershed (mines/square km)	StreamCat
DepNWs	Mean atmospheric nitrogen deposition in watershed 2018	National Atmospheric Deposition Program
DepSWs	Mean atmospheric sulfur deposition in watershed 2018	National Atmospheric Deposition Program
MineDensWs	Density of coal mines sites within watershed (mines/square km)	StreamCat
N_inputs	Sum of anthropogenic nitrogen inputs: N_Fert_FarmWs, N_Fert_UrbanWs, N_Human_WasteWs, N_Livestock_WasteWs	Sabo et al. 2023
P_inputs	Sum of anthropogenic phosphorus inputs: P_fertilizerWs, P_human_wasteWs, P_livestock_WasteWs, P_nf_fertilizerWs	Sabo et al. 2023
PctCropWs	% of watershed area classified as crop land use	StreamCat
PctCropWsRp100	% of watershed riparian area classified as crop land use	StreamCat
PctNatWs	% of watershed with natural vegetation cover	StreamCat
TPRCP	Ln(Mean total annual precipitation, 2018 or 2019)	PRISM
RdDensWs	Density of roads (2010 Census Tiger Lines) within watershed (km/square km)	StreamCat
MSAT	Mean summer (July/August) air temperature (C°), 2018 or 2019	PRISM
W1_HAG	Agricultural disturbance adjacent to a stream	NRSA Field Data
NABD_NrmStorWs	Volume all reservoirs per unit area of watershed (cubic meters/square km)	StreamCat

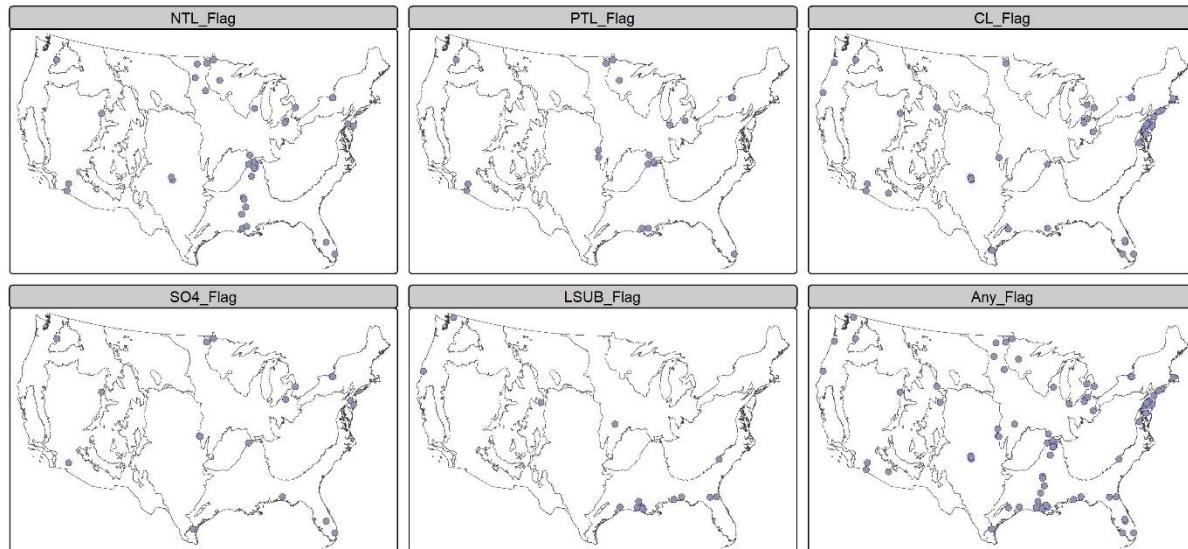
32 Table S2: Variable importance rankings for random forest models and values used to estimate
 33 physiochemical conditions in the absence of anthropogenic disturbance. The 5 most important
 34 variables for each model are bold.

Predictor Variable	NTL	PTL	CL	SO4	SUBD	Range of Predictor Present-day	Value Without Disturbance
Geoclimatic							
BFIWs	6	13	7	8	-	2.6 - 88	--
ClayWs	-	10	8	12	-	2.35 - 60.6	--
ElevWs	10	4	6	9	-	2.02 - 3587.98	--
KffactWs	13	11	15	14	9	0.01 - 0.48	--
DepCIWs	-	-	9	-	-	0.13-43.16	--
DepPWs	-	5	-	-	-	2.03-15.9	--
NWs	9	-	11	4	-	0-1.59	--
P2O5Ws	16	9	18	15	-	0.02-2.42	--
PctAlluvCoastWs	-	-	-	-	6	0-100	--
PctHbWetWs	7	8	10	16	-	0-40.35	--
PermWs	14	14	17	13	-	0.32-45.23	--
RunoffWs	2	2	4	1	2	0.05-3731.74	--
SandWs	-	-	-	-	8	2.8-92.04	--
StreamPower	-	-	-	-	1	0-0.44	--
SWs	12	15	5	3	-	0.01-23.66	--
WsAreaSqKm	15	12	14	11	5	0.21-2,874,021.46	--
Anthropogenic							
CoalMineDensWs	-	-	19	7	13	0 - 5.74	0
DepNWs	5	-	-	-	-	1.12 - 26.15	5 kgN/ha/yr*
DepSWs	-	-	12	10	-	0.22 - 5.88	1.65 kgS/ha/yr*
MineDensWs	-	-	16	18	12	0 - 0.13	0
N_inputs	3	-	-	-	-	0 - 6.22	0
P_inputs	-	7	-	-	-	0 - 2.19	0
PctCropWs	1	1	3	2	-	0 - 96.7	0
PctCropWsRp100	-	-	-	-	4	0 - 95.52	0
PctNatWs	-	-	-	-	3	0 - 100	100
TPRCP	11	16	13	17	-	4.61 - 8.28	Average 1900-1950
RdDensWs	4	6	2	5	10	0 - 15.15	0
MSAT	8	3	1	6	-	11.71 - 36.11	Average 1900-1950
W1_HAG	-	-	-	-	7	0 - 2.18	0
NABD_NrmStorWs	-	-	-	-	11	0 - 1,886,390.68	0

35 * Estimates for atmospheric nitrogen and sulfur deposition were obtained from Clark et al. 2018.

36 Although they used 0.4 kgN/ha and 0.1kgS/ha no sites had deposition values below this level.

37 Instead, we selected the higher values of 3-5KgN/ha as reasonable estimates for background
 38 deposition values before 1900. For S deposition, other efforts have estimated pre-industrial S
 39 deposition at 0.32–2.98 kgS/ha (Granat et al. 1976, Fakhraei et al. 2016) and we used the used
 40 middle number of this interval as a reasonable estimate for background.

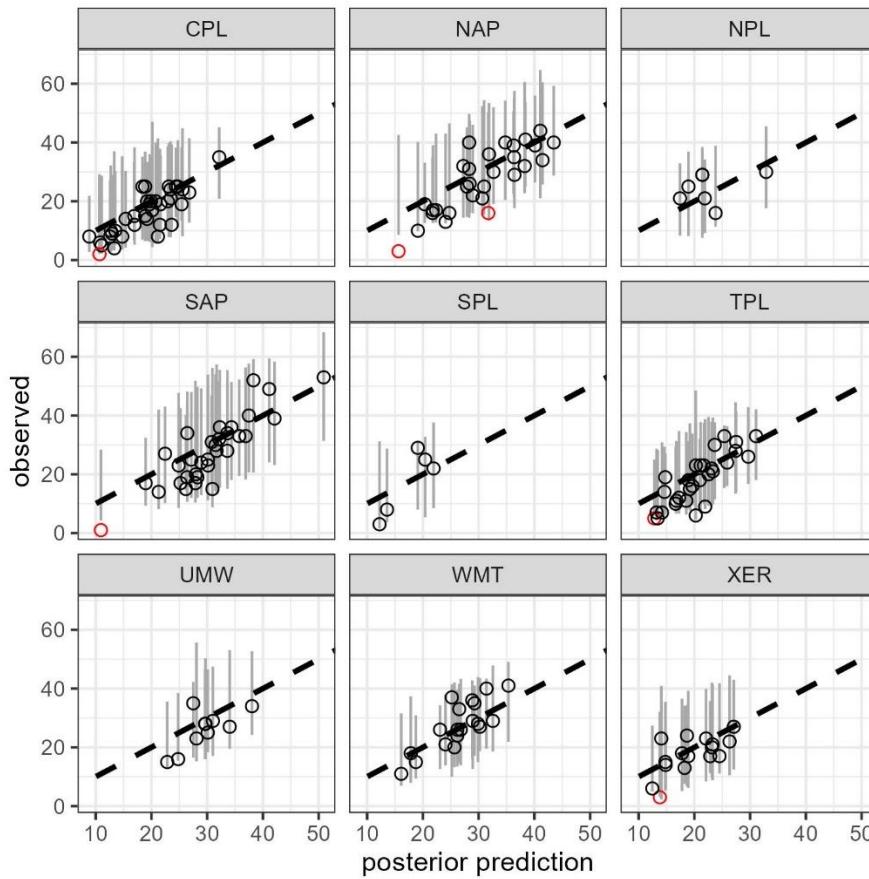


41
42 Figure S1: Sites flagged for potential extrapolation after removing anthropogenic effects from
43 the random forest models. NTL = total nitrogen; PTL is total phosphorus; CL is chloride; SO4 is
44 sulfate; LSUB = substrate diameter, Any = sites that were flagged for extrapolation for any
45 model and removed from the analysis. The number of sites removed is as follows (CPL = 40
46 (18%); NAP = 2 (1%); NPL = 1 (1%); SAP = 1 (<1%); SPL = 6 (3%); TPL = 13 (6%); UMW=3
47 (1%); WMT = 2(1%) and XER = 5 (3%))
48
49

50 Table S3. Variance inflation factors (VIF) for predictor variables used for joint species
51 distribution models. In general, VIF >5 indicates a strong correlation between variables while
52 values VIF < 3 indicate low correlation between values. Bold font = VIF >3; NTL = Total
53 Nitrogen, PTL = Total Phosphorus, CL = Chloride, SO4 = Sulfate, SUBD = Substrate Diameter,
54 RPDI = Riparian Disturbance index, MSAT = Mean Summer Air Temperature, PRCP = Total
55 Annual Precipitation.
56

	NTL (ug/L)	PTL (ug/L)	CL (mg/L)	SO4 (mg/L)	SUB (mm)	RPDI	MSAT (DegC)	TPRCP (mm)
CPL	2.17	1.87	2.83	2.5	1.09	1.3	1.25	1.14
NAP	3.08	2.6	3.34	1.67	1.1	1.28	2.47	1.07
NPL	2.8	2.2	2.82	2.94	1.39	1.11	1.84	1.5
SAP	2.42	1.85	2.6	1.49	1.13	1.3	1.21	1.04
SPL	2.49	2.57	2.65	2.12	1.22	1.17	1.63	1.28
TPL	1.57	1.35	1.22	1.45	1.11	1.11	1.52	1.5
UMW	2.37	1.92	3.13	2.57	1.09	1.46	4.25	1.15
WMT	1.79	1.6	2.22	1.86	1.25	1.34	1.47	1.29
XER	2.28	1.77	4.1	3.23	1.29	1.19	1.69	1.44

57
58
59



60
61
62 Figure S2. Validation of JSDM using revisit surveys. As part of NRSA, a subsample of sites are
63 resurveyed. We used these data to provide additional validation of the predictive capabilities of JSDMs.
64 We found that when using condition prediction (Ovaskainen and Abrego 2020), the posterior
65 distribution of all JSDM typically contained all observed genus richness values. This result suggests that
66 the observed genus richness could be a random sample drawn from our modeled posterior distribution
67 and suggests that our models can accurately predict the occurrence of taxa during the initial visit. Black
68 points indicate that the observed genus richness is within the predicted posterior distribution; Red
69 points indicate that the observed genus richness is outside the predicted posterior distribution; Dotted
70 line is 1:1 between observed and posterior mean richness and grey lines represent the distribution of
71 3,000 samples from the posterior distribution.

72 Table S4. Regional means of present-day (PD) and hindcasted (HC) values for each environmental gradient. Parentheses are 10th and
 73 90th quantiles of observed hindcasted values. NTL = Total Nitrogen, PTL = Total Phosphorus, CL = Chloride, SO4 = Sulfate, SUBD =
 74 Substrate Diameter, RPDI = Riparian Disturbance index, MSAT = Mean Summer Air Temperature, PRCP = Total Annual
 75 Precipitation.

	NTL (µg/L)		PTL (µg/L)		CL (mg/L)		SO4 (mg/L)		SUB (mm)		RPDI		MSAT (DegC)		TPRCP (mm)	
	PD	HC	PD	HC	PD	HC	PD	HC	PD	HC	PD	HC	PD	HC	PD	HC
CPL	702.5 (299.2 - 2027.2)	380 (315.4 - 566.4)	77.82 (23.07 - 234.27)	48.67 (25.67 - 77.47)	8.7 (2.89 - 68.94)	3.7 (2.92 - 7.65)	6.3 (1.08 - 34.55)	3.29 (1.82 - 11.02)	0.35 (0.01 - 2.73)	3.32 (0.7 - 14.32)	0.44 (0.03 - 1.58)	0.33 (0.03 - 0.33)	27.49 (25.63 - 28.54)	27.02 (24.69 - 27.84)	1623.25 (1178.37 - 1990.68)	1219.49 (1015.53 - 1664.18)
NAP	422 (187- 1301)	286 (188- 327)	23.62 (7.25- 71.81)	17.76 (10.04- 27.89)	16.58 (0.46- 72.61)	1.92 (0.3- 3.82)	5.57 (1.71- 15.63)	2.9 (1.74- 6.07)	12.09 (0.09- 122.43)	14.36 (3.4- 47.45)	0.65 (0- 2.02)	0.33 (0- 0.33)	21.49 (18.99 - 23.67)	20.41 (18.31 - 21.75)	1299.05 (1106.82- 1629.65)	1089.44 (949.5- 1234.19)
NPL	881.5 (265.9 - 2193.5)	462.75 (216- 694)	85.89 (18.01 - 366.21)	67.33 (38.47- 85.9)	12.09 (2.34- 50.02)	10.16 (1.45- 19.62)	486.24 (18.73- 1340.1)	172.31 (9.75- 517.88)	0.51 (0.01- 18.45)	3.1 (1.19- 29.75)	1.27 (0.38- 2.08)	0.33 (0.33- 0.33)	20.52 (18.34 - 22.49)	21.19 (19.1- 23.77)	437.42 (320.65- 632.12)	361.68 (282.17- 428.71)
SAP	556 (173- 1876)	284 (158- 350)	29.26 (9.83- 128.25)	24.3 (16.41- 45.85)	5.82 (1.34- 47.53)	2.7 (1.08- 3.82)	7.47 (1.55- 80.36)	3.88 (1.88- 8.95)	16.13 (0.35- 206.69)	14.32 (6.1- 57.76)	0.77 (0.05- 2.29)	0.33 (0.05- 0.33)	24.76 (22.09 - 26.68)	24.05 (21.37 - 26.37)	1476.55 (1108.36- 1809.43)	1110.8 (920.87- 1283.9)
SPL	1165 (438.5 - 2961)	405.25 (301.75 - 627)	143.62 (26.05 - 774.69)	56.88 (37.68- 76.58)	29.21 (4.51- 358.38)	10.47 (4.29- 29.42)	86.1 (7.94- 898.01)	45.6 (8.5- 334.85)	0.35 (0.02- 19.61)	3.18 (0.92- 12.68)	1.02 (0.21- 2.29)	0.33 (0.21- 0.33)	25.84 (21.9- 29.66)	26.17 (22.03 - 28.92)	616.17 (393.76- 1051.75)	534.19 (312.86- 977.73)
TPL	1868 (647.6 - 9147.8)	418 (330.6- 578)	165.08 (54.39 - 444.97)	50.96 (40.55- 63)	18.29 (7.1- 62.09)	5.7 (3.7- 10.38)	32.99 (11.93- 256.93)	13.48 (5.28- 105.55)	0.35 (0.01- 16.28)	7.79 (2.51- 18.96)	0.83 (0.13- 2.29)	0.33 (0.13- 0.33)	23.36 (21.05 - 25.28)	23.63 (21.3- 25.38)	1004.99 (678.76- 1395.29)	775.15 (537.81- 922.38)
UMW	1188 (352.2 - 5320)	371 (289.6- 511.49)	57.4 (16.93 - 166.88)	36.03 (17.9- 53.25)	9.98 (0.76- 29.9)	2.72 (0.3- 5.21)	9.28 (1.12- 40.68)	3.65 (1.34- 8.36)	0.35 (0.01- 12.6)	6.1 (2.19- 27.43)	0.62 (0- 2.1)	0.33 (0- 0.33)	20.84 (18.13 - 22.56)	20.98 (18.42 - 22.38)	961.61 (801.29- 1395.23)	799.34 (623.5- 877.62)
WMT	133 (50- 385.6)	133 (66- 235.1)	22.12 (6.97- 82.29)	21.14 (9.3- 53.81)	0.92 (0.12- 9.14)	0.52 (0.16- 3.13)	2.97 (0.47- 24.57)	3.2 (0.86- 15.04)	38.14 (0.97- 192.14)	38.17 (6.1- 99.8)	0.48 (0- 1.73)	0.33 (0- 0.33)	18.08 (13.97 - 22.33)	16.96 (12.57 - 21.33)	648.86 (349.11- 1670.25)	688.27 (386.2- 1778.26)

XER	333 (107.2 - 1245.6)	253 (133.8- 418)	53.87 (14.3- 240.71)	42.76 (15.83- 84.65)	5.23 (0.79- 132.36)	3.17 (0.68- 20.77)	25.08 (1.54- 324.3)	20.85 (1.93- 112.43)	2.18 (0.02- 66.09)	10.78 (2.1- 44.24)	1.14 (0.11- 2.54)	0.33 (0.11- 0.33)	22.76 (18.81 - 29.64)	21.33 (17.46 - 26.95)	287.58 (170.89- 511.45)	337.66 (223.14- 702.18)
-----	-----------------------------------	------------------------	--------------------------------	----------------------------	---------------------------	--------------------------	---------------------------	--------------------------------	--------------------------	--------------------------	-------------------------	-------------------------	--------------------------------	--------------------------------	-------------------------------	-------------------------------

For Review Only

77

78

For Review Only

1 A model-based assessment of anthropogenic disturbance on lotic macroinvertebrate assemblages

2

3 Darin A. Kopp^{1,2*}, John L. Stoddard²Stoddard¹, Philip R. Kaufmann^{2,3}Kaufmann^{1,2}, Alan T.
4 Herlihy³Herlihy², Ryan Hill²A. Hill¹, Meredith Brehob⁵M. Brehob³, Robert Sabo⁶D Sabo⁴

5

6 ¹~~Oak Ridge Institute for Science and Education Participant c/o US EPA, ORD, Corvallis,~~
7 ~~Oregon~~

8 ²~~US~~¹US EPA, Office of Research and Development, Corvallis, Oregon, USA

9 ³~~Department~~²Department of Fisheries, Wildlife, & Conservation Sciences, Oregon State
10 University,

11 Corvallis, Oregon, USA

12 ⁴~~Current affiliation: Oregon Department of Fish and Wildlife, Salem, Oregon~~

13 ⁵~~Oak~~³Oak Ridge Institute for Science and Education Participant c/o US EPA, ORD, Washington,
14 D.C., USA

15 ⁶~~US~~⁴US EPA, Office of Research and Development, Center for Public Health and Environmental
16 Assessment, ~~DC~~, Washington, DC, USA

17

18 *Correspondence: kopp.darin@epa.gov

19

20
21 Keywords: Joint Species Distribution Modeling, Richness, National Rivers and Streams
22 Assessment, Benthic Macroinvertebrate, Reference Sites, National Aquatic Resource Surveys.

23

24

25 [Open research statement: Benthic macroinvertebrate occurrence data and physiochemical](https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys)
26 [variables were retrieved from the 2018-2019 National Rivers and Stream Assessment](https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys)
27 (<https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys>)
28 Benthic macroinvertebrate data can be downloaded as .csv file from NRSA 1819 Benthic
29 Macroinvertebrate Count - Data (CSV) (csv). Physiochemical data can be downloaded as .csv
30 files from NRSA 1819 Water Chemistry_CHLA - Data (CSV) (csv) and NRSA 1819 Physical
31 Habitat Larger Set of Metrics - Data (CSV) (csv) Landscape variables were retrieved from
32 (https://www.epa.gov/national-aquatic-resource-surveys/streamcat-metrics-and-definitions_and
33 PRISM Climate Group <https://prism.oregonstate.edu/>

34

35 Abstract:

36 Traditionally, the effects of anthropogenic disturbance on biological assemblages are
37 elucidated by comparing an assemblage observed at a site to one that represents a minimally
38 disturbed state. Unfortunately, defining a minimally disturbed state is extremely challenging
39 because of the extent of human disturbance. We use a national scale dataset and a two-stage
40 [modelingmodel-based](#) approach to assess how benthic macroinvertebrate assemblages at
41 1,[824748](#) sites would change if common anthropogenic disturbances were removed from in-
42 stream physiochemical variables. First, we used random forest models and current landscape data
43 to predict physiochemical conditions and then infer abiotic condition in the absence of
44 disturbance. Second, we combined these estimates with joint species distribution models to
45 predict the assemblage that would occur in these undisturbed conditions. Random forest models
46 explained 48 – 75% of the variation in total nitrogen, phosphorous, sulfate, chloride, and
47 substrate diameter. Generally, nutrient and salinity concentrations were higher, and substrates
48 were finer than predicted to be without disturbances. Using this physiochemical data, joint
49 species distribution models accurately explained genus richness ($R^2 = 0.73 – 0.85$) and
50 composition (Jaccard similarity index = 0.48 – 0.55). Depending on the ecoregion, we found that
51 genus richness could change at 26 – 61% of sites if disturbance was removed. Different

52 responses were observed for insect and non-insect taxa. For example, under anthropogenic
53 disturbance, occurrence probabilities for Ephemeroptera, Plecoptera and Trichoptera tended to
54 decrease at 5 – 26% of sites while occurrence probabilities for Mollusca and other non-insect,
55 non-arthropod taxa increased at 5 – 33% and 11 – 24% of sites, respectively. Importantly, our
56 framework provides an avenue for evaluating the effects of anthropogenic disturbance on
57 macroinvertebrate assemblages without ~~relying on defining identifying~~ reference sites.

58

59 Introduction:

60 Quantifying the extent and magnitude of anthropogenic disturbance on ecosystems
61 requires a benchmark that represents a desired, expected, or previous condition (~~Stoddard et al.~~
62 ~~2006, Hawkins et al. 2010, McNellie et al. 2020~~)~~(Stoddard et al. 2006, Hawkins et al. 2010,~~
63 ~~McNellie et al. 2020)~~. Ideally, benchmarks should represent chemical, physical, and biological
64 characteristics with minimal anthropogenic disturbance (~~Reynoldson et al. 1997, Stoddard et al.~~
65 ~~2006~~)~~(Reynoldson et al. 1997, Stoddard et al. 2006)~~. However, for many ecosystems the
66 minimally disturbed condition is difficult to quantify or may no longer exist (~~Dudgeon et al.~~
67 ~~2006, Vörösmarty et al. 2010, McNellie et al. 2020~~)~~(Dudgeon et al. 2006, Vörösmarty et al.~~
68 ~~2010, McNellie et al. 2020)~~. The concept of “least disturbed” describes conditions that are
69 derived from a collection of sites that represent the best available, given today’s landscape
70 (~~Stoddard et al. 2006~~)~~(Stoddard et al. 2006)~~. Unfortunately, because anthropogenic disturbance is
71 not evenly distributed across the landscape, sites in least disturbed condition can be spatially
72 ~~aggregated biased and~~/or vary in quality ~~among ecoregions~~ (~~Herlihy et al. 2008, McNellie et al.~~
73 ~~2020~~)~~(Herlihy et al. 2008, McNellie et al. 2020, Yuan et al. 2024)~~. This makes it challenging to
74 ~~elucidate the effects of anthropogenic disturbance on biological assemblages because the~~

75 condition represented by a suite of. If least disturbed sites are spatially biased, their
76 ecological, chemical, and physical conditions may be unattainable for some locations that need
77 to be assessed because of differences in natural setting or biogeographic history.
78 Alternatively, if reference site quality is lower for regions that have had some ecosystems
79 because of a relatively long legacy of human intensification (McNellie et al. 2020), some sites
80 will be judged against a lower benchmark and erroneously appear to be in better condition
81 relative to others. Although least disturbed reference sites provide reasonable and defensible
82 benchmarks for assessing biological condition (Stoddard et al. 2006, Herlihy et al. 2008,
83 Mitchell et al. 2025), they are typically not undisturbed, nor evenly distributed among all
84 ecosystem types (McNellie et al. 2020, Yuan et al. 2024). Novel approaches and concepts are
85 urgently needed to overcome these limitations.

86 Quantifying taxon-environment relationships for all taxa within a region with empirical
87 models may provide an alternative to relying on reference sites because they elucidate how
88 individual taxa and entire assemblages change along physiochemical gradients (Chessman and
89 Royal 2004, Kilgour and Stanfield 2005, Elias et al. 2016, Kopp et al. 2023). Because
90 anthropogenic disturbance typically manifests as altered physiochemical conditions (Tang et al.
91 2020, Fergus et al. 2023), quantifying potential environmental diversity across space, could be
92 relevant for understanding how organisms may change with disturbance across time (Blüthgen et
93 al. 2022). For example, taxa that have a negative relationship along an environmental gradient
94 could be expected to become less prevalent if human disturbance contributes to more extreme
95 environmental conditions and more prevalent if reducing human disturbance contributes to less
96 extreme conditions (Kopp et al. 2023). From these relationships it is possible to predict which
97 taxa from the regional species pool could occur at a site if it were minimally disturbed and

98 compare them to the taxa that were actually observed (Chessman and Royal 2004, Elias et al.
99 2016). If the minimally disturbed assemblage is similar to the observed assemblage, then the site
100 may be considered unaffected by anthropogenic disturbances.

101 Advances in species distribution modeling enhance our ability to quantify taxon-
102 environment relationships (Franklin 2010, Guisan et al. 2017, Ovaskainen and Abrego 2020).

103 Species distribution modeling consists of a diverse set of methods that typically focus on
104 quantifying taxon-environment relationships for individual species (Elith and Leathwick 2009).

105 Joint species distribution modeling (JSDM) is a recent multivariate extension of single species
106 distribution modeling, capable of estimating taxon-environment relationships for all members of
107 an assemblage simultaneously. Because of their complex structure, JSDMs are often fitted using
108 Bayesian inference. As such, the resulting posterior distribution can be used for hypothesis
109 testing (Johnson et al. 2022). For example, given a posterior distribution of values representing a
110 biological assemblage (e.g. taxonomic richness) in the absence of human disturbance, it is
111 possible to evaluate whether an observed assemblage is consistent with this distribution. In
112 addition, these models use latent factors to account for unmeasured environmental variables and
113 potential biotic interactions (Warton et al. 2015, Ovaskainen et al. 2017, Ovaskainen and Abrego
114 2020). This feature is particularly important for biological assessments because it can strengthen
115 the inference gained from evaluating a specific suite of environmental variables that are typically
116 altered by human disturbance. Thus, the ability to model all taxa in an assemblage and
117 probabilistically evaluate whether an observation is consistent with a posterior distribution
118 makes JSDMs well suited for understanding how anthropogenic factors affect biological
119 assemblages.

120 Although taxon-environment relationships are necessary to circumvent the need for
121 reference sites, they require a concurrent expectation of undisturbed physiochemical conditions
122 to fully understand the effects of anthropogenic activities. Traditionally, undisturbed
123 physiochemical conditions are obtained, or modeled, from least-disturbed reference sites (Olson
124 and Hawkins 2012, Olson and Cormier 2019). However, empirical models have also linked
125 changes in the physiochemical environment to anthropogenic activities and been used to estimate
126 conditions if anthropogenic disturbance is reduced or eliminated without reference sites (Dodds
127 and Oakes 2004, Herlihy and Sifneos 2008, Soranno et al. 2011). Random forest models are a
128 machine learning algorithm that have superior predictive performance compared to other
129 modeling techniques (Prasad et al. 2006, Peters et al. 2007) and are often used to model water
130 quality parameters and physical habitat characteristics (Olson and Hawkins 2013, Olson and
131 Cormier 2019, Sabo et al. 2023, Yuan et al. 2024). Further these models are more robust to
132 nonlinearities, multicollinearity, and overfitting (Breiman 2001) and can evaluate the relative
133 importance of different predictor variables (Lin et al. 2021, Sabo et al. 2023). Because these
134 algorithms can accommodate complex relationships and a relatively high number of covariates,
135 they are well suited for evaluating the relative importance of natural and anthropogenic variables
136 on physiochemical conditions and, in turn, predicting the abiotic conditions if anthropogenic
137 disturbances were removed (Yuan et al. 2024). Indeed, efforts that combined estimates of
138 undisturbed physiochemical conditions with biological models have made progress towards
139 addressing the limitations of reference site-based approaches (Chessman and Royal 2004,
140 Kilgour and Stanfield 2005, Elias et al. 2016, Yuan et al. 2024).

141 In rivers and streams, ~~preserving and maintaining biological integrity is a central~~
142 ~~management objective (Jackson and Davis 1994, Paulsen et al. 2008, Hering et al. 2010).~~

143 ~~Benthic macroinvertebrate assemblages are often surveyed as part of assessing biological~~
144 ~~integrity assessments of biological integrity often rely on surveys of benthic macroinvertebrate~~
145 ~~assemblages~~ because they are diverse, relatively easy to sample, and respond to changes in the
146 physiochemical environment associated with anthropogenic disturbance (Hughes and Peck 2008,
147 Buss et al. 2014)(Hughes and Peck 2008, Buss et al. 2014). To elucidate the effects of
148 anthropogenic activities on biotic integrity, benthic macroinvertebrate assemblages observed at a
149 site are typically compared to reference assemblages collected from least disturbed sites
150 (~~Hawkins 2006, Stoddard et al. 2008~~)(Hawkins 2006, Stoddard et al. 2008). Reference
151 ~~assemblages should be obtained from sites that are minimally disturbed (Reynoldson et al. 1997)~~
152 ~~but, given the pervasiveness of anthropogenic disturbance and its covariance with geoclimatic~~
153 ~~predictors, this is often unattainable. Instead, macroinvertebrate assemblages are typically~~
154 ~~collected from the best available sites such that benchmarks typically represent a least disturbed~~
155 ~~condition (Yuan 2006, Herlihy et al. 2008, Chen et al. 2019)~~. Although least disturbed
156 ~~benchmarks provide a reasonable and defensible management target, they are not undisturbed.~~
157 ~~Thus, they can mask the true extent of anthropogenic disturbance.~~

158 ~~Aquatic organisms are directly affected by the physiochemical environment of the stream~~
159 ~~which, in turn, is determined by a combination of current and past geoclimatic and anthropogenic~~
160 ~~factors in the watershed (Tang et al. 2020, Fergus et al. 2023)~~. Empirical models that link the
161 ~~physiochemical environment to anthropogenic disturbance in the watershed could be used to~~
162 ~~estimate physiochemical characteristics if anthropogenic disturbance is reduced (Dodds and~~
163 ~~Oakes 2004, Herlihy and Sifneos 2008, Soranno et al. 2011)~~. These estimates can then be
164 ~~combined with taxon–environment relationships to elucidate how individual taxa and entire~~
165 ~~assemblages might respond to changing physiochemical conditions (Chessman and Royal 2004,~~

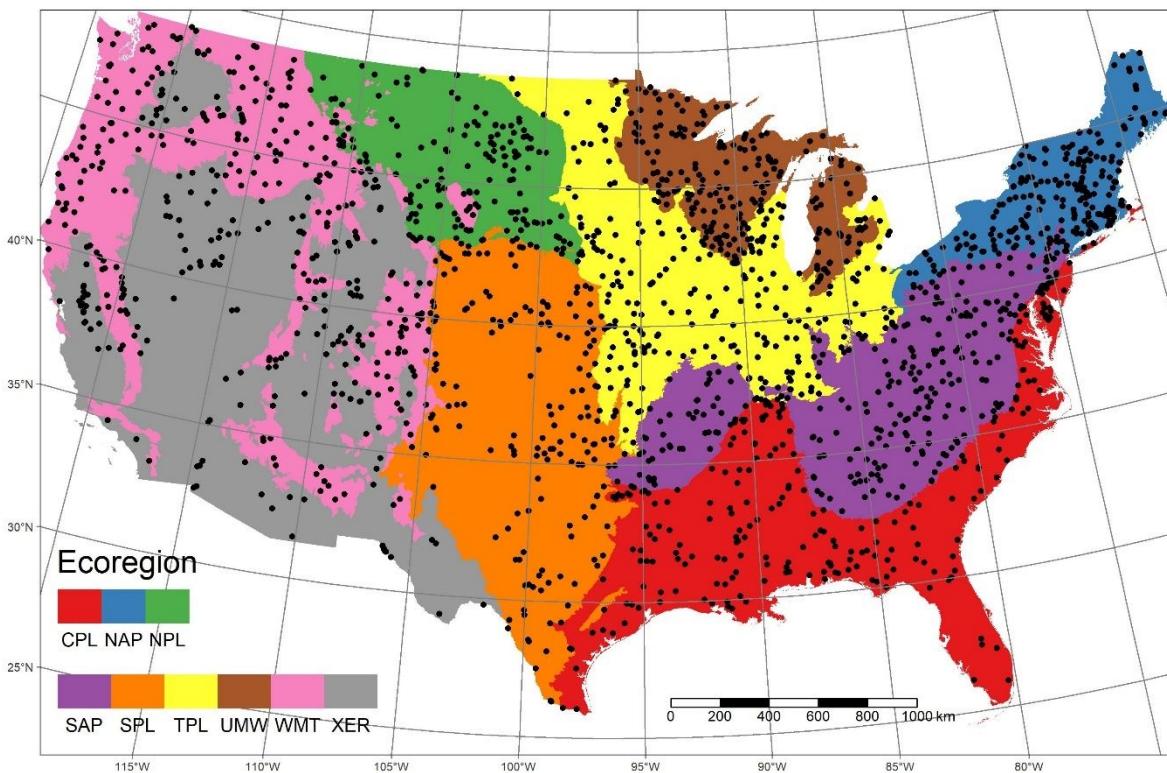
166 Kilgour and Stanfield 2005, Elias et al. 2016, Kopp et al. 2023). Because physiochemical
167 characteristics are often used as management targets, this framework may offer a more direct
168 link between management intervention and biotic responses compared to using reference sites
169 (Hering et al. 2010, McNellie et al. 2020, Deflem et al. 2021).

170 Here, we use a two-stage modeling approach to assess how benthic macroinvertebrate
171 assemblages at 1,824. Here, we develop a model-based assessment as a complementary method
172 to evaluate how benthic macroinvertebrate assemblages at 1,748 stream sites may change if the
173 effects of major anthropogenic disturbances were removed from the physiochemical
174 environment. First, we use random forest models to relate in-stream physiochemical conditions
175 to a suite of geoclimatic and anthropogenic factors and used these models to infer abiotic
176 conditions if anthropogenic disturbance was eliminated. Second, we combined these estimates
177 with joint species distribution models (JSDMs) to predict how biological assemblages
178 could change if physiochemical conditions were not altered by anthropogenic disturbance. We
179 focus on Because JSDMs can leverage latent factors to account for unmeasured environmental
180 factors, including potential biotic interactions (Ovaskainen et al. 2016), we focus exclusively on
181 environmental variables that are commonly associated with anthropogenic disturbances.
182 Specifically, we evaluate the effects of changing nutrient and salinity concentrations, physical
183 habitat, and climate on benthic macroinvertebrate assemblages because these gradients are
184 known to be influenced by humans commonly used as abiotic screens used to identify least
185 disturbed reference sites (Herlihy et al. 2008, Paulsen et al. 2008, USEPA 2023)(Herlihy et al.
186 2008, Paulsen et al. 2008, USEPA 2023, Mitchell et al. 2025) and affect benthic
187 macroinvertebrate distributions (Kopp et al. 2023)(Kopp et al. 2023).

188

189 Methods:

190 For our analysis, we focused on biological and environmental data collected at sites
191 distributed among 9 ecoregions (Figure 1) and surveyed during the 2018/2019 National Rivers
192 and Streams Assessment (NRSA). NRSA is a collaboration between the United States
193 Environmental Protection Agency (USEPA) and state, tribal, and federal partners to assess the
194 chemical, physical, and biological condition of streams and rivers of the United States
195 (<https://www.epa.gov/national-aquatic-resource-surveys/nrsa>). Every five years, beginning in
196 2008/2009, NRSA ~~uses a statistical survey design to randomly selectsurveys~~ ~2,000 stream/river
197 locations that are ~~representative of the entire population of streams and rivers in the contiguous~~
198 ~~US (Olsen and Peck 2008).selected using a probabilistic survey design and handpicked by state~~
199 ~~and tribal partners.~~ These sites are surveyed for biological, chemical, and physical characteristics
200 using standardized field protocols ~~(USEPA 2023). Only sites included in the probabilistic sample~~
201 ~~are selected to be representative of the entire population of streams and rivers in the contiguous~~
202 ~~US and used for assessment (Olsen and Peck 2008). In addition, approximately 10% of the~~
203 ~~probabilistic sites are revisited to assess within-year variability and approximately 30% of these~~
204 ~~sites are resampled in the next survey cycle to evaluate among-year variability.~~



205
206 **Figure 1: Ecoregions and survey locations for the National Rivers and Streams Assessment**
207 **2018–2019 survey. CPL = Coastal Plains, NAP = Northern Appalachians, NPL = Northern**
208 **Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL = Temperate Plains, UMW =**
209 **Upper Midwest; WMT = Western Mountains; XER = Xeric**

210

211 *Benthic macroinvertebrate assemblages and physiochemical variables*

212 Procedures for collecting benthic macroinvertebrates are described in detail elsewhere

213 ([Hughes and Peck 2008](#), [USEPA 2017a](#)) ([Hughes and Peck 2008](#), [USEPA 2017a](#)). Briefly, each

214 survey location was defined as a length of stream or river equal to a multiple of its channel

215 width. In wadeable and boatable sites <13 m wide, the reach length sampled was equal to 40

216 channel widths, or a minimum of 150 m. For wadeable sites >13 m wide, and boatable sites, the

217 minimum reach length sampled was the longer, or a maximum of 500 m or 20 channel

218 widths4km. Macroinvertebrate samples were collected along 11 equally spaced cross-section

219 transects along the reach. In wadeable sites, samples were collected in an alternating left, center,

220 right order along the transects using a D-frame kick net (500-um mesh, 0.09 m² area). In

221 boatable sites, samples were collected along the left or right wadeable margin from a 1m linear
222 sweep of the dominant habitat using a D-frame kick net (500-um mesh, 0.3048 m² area).
223 Samples from each survey location were combined into a single composite sample, preserved in
224 ethanol, and sent to a taxonomist for subsampling and identification. ~~We used presence/absence~~
225 ~~data from the ~2,000 benthic macroinvertebrate collections and focused on taxa that were~~
226 ~~collected from a single site visit, identified to genus, and occurred at ≥10% of sites within an~~
227 ~~ecoregion (Table 1).~~

228 ~~We~~To assess the effects of anthropogenic disturbance on in-stream physiochemical
229 ~~conditions and in turn macroinvertebrate assemblages, we~~ focused on in situ nutrient
230 concentrations, salinity, and physical habitat variables,~~because they.~~ These environmental
231 ~~gradients~~ are commonly altered by anthropogenic activities in the surrounding watershed ~~and~~
232 ~~frequently used as abiotic screens to identify least-disturbed reference sites (Herlihy et al. 2008,~~
233 ~~Paulsen et al. 2008, USEPA 2023)(Herlihy et al. 2008, Paulsen et al. 2008, USEPA 2023) and~~
234 ~~influence macroinvertebrate distributions (Kopp et al. 2023). We selected total nitrogen (NTL)~~
235 ~~and total phosphorus (PTL) as indicators of nutrients; chloride (CL) and sulfate (SO4) as~~
236 ~~indicators of salinity; and mean substrate diameter (SUBD) and riparian disturbance index~~
237 ~~(RPDI) as indicators of physical habitat. A. Specifically, we selected total nitrogen (NTL) and~~
238 ~~total phosphorus (PTL) as indicators of excess nutrients supplied from agriculture and/or~~
239 ~~urbanization (Herlihy et al. 1998) and chloride (CL) and sulfate (SO4) as indicators of salinity.~~
240 ~~Elevated CL is a general indicator of human disturbance in the catchment (Herlihy et al. 1998)~~
241 ~~and SO4 can indicate sites that are affected by mine drainage (Herlihy et al. 1991). Mean~~
242 ~~substrate diameter (SUBD) and riparian disturbance index (RPDI) were selected as indicators of~~
243 ~~physical habitat because human activities can increase fine sediment inputs or directly modify~~

244 the riparian area (Kaufmann 1999, Kaufmann et al. 2022b). In addition, we included mean
245 summer air temperature (Mean July and August temperature, MSAT) and total annual
246 precipitation (TPRCP) to assess potential changes associated with differences from a 1900-1950
247 baseline. Although other factors are affected by human disturbances (e.g., metals and pesticides
248 and hydrologic alteration), these variables provide a reasonably broad characterization of the
249 physiochemical environment at a site (Herlihy et al. 2008, Paulsen et al. 2008)

250 During the survey, a single water sample was collected at each site and shipped to a
251 central analytical laboratory. PTL and NTL were measured by persulfate digestion and
252 colorimetry and CL and SO₄ were measured by ion chromatography (USEPA 2017b)(USEPA
253 2017b). SUBD is the geometric mean of the numeric value assigned to substrate size classes
254 measured in the field (Kaufmann 1999).(Kaufmann 1999). RPDI summarizes the
255 presence/absence of 11 categories of human disturbance, including buildings, landfill/ trash,
256 logging, mining, developed parks or lawns, pavement or cleared lots, pipes (withdrawal or
257 wastewater), roads, row crops, pastures or hayfields, and walls or revetments in the riparian area,
258 adjacent to each to the 11 transects where macroinvertebrates were collected (Kaufmann 1999,
259 USEPA 2017a, Kaufmann et al. 2022)(Kaufmann 1999, USEPA 2017a, Kaufmann et al. 2022b).
260 In addition, we included mean summer air temperature (Mean July and August temperature,
261 MSAT) and total annual precipitation (TPRCP) for either 2018 or 2019 (depending on sample
262 year) obtained from the PRISM datasetClimate Group (<https://prism.oregonstate.edu/>).

263

264 Table 1: Sites, genera, and environmental variables included in the analysis. Values for the environmental variables are median values
 265 for each region. Range is given in parentheses. NTL = Total Nitrogen, PTL = Total Phosphorous, CL = Chloride, SO₄ = Sulfate, RPDI
 266 = Riparian Disturbance Index, SUBD = Substrate Diameter, TPRCP = Total Precipitation and MSAT = Mean Summer Air
 267 Temperature. CPL = Coastal Plains, NAP = Northern Appalachians, NPL = Northern Plains, SAP = Southern Appalachians, SPL =
 268 Southern Plains, TPL = Temperate Plains, UMW = Upper Midwest; WMT = Western Mountains; XER = Xeric.
 269

Region	Site (#)	Genera (#)	NTL (ug/L)	PTL (ug/L)	CL (mg/L)	SO ₄ (mg/L)	RPDI	SUBD (mm)	TPRCP (mm)	MSAT (°C)
CPL	226	71	786 (91-7713)	84.81 (5.2-3922.25)	10.06 (0.52-4797.9)	7.83 (0.11-1900.89)	0.5 (0.5-39)	0.35 (0.01-341.65)	1613.83 (388.66-2397.1)	27.46 (23.75-31.68)
NAP	228	127	431 (81-6413)	23.76 (3.12-587.2)	16.73 (0.07-668.16)	5.58 (0.06-467.93)	0.65 (0.5-24)	11.66 (0.01-864.9)	1295.44 (888.91-1883.17)	21.49 (17.24-24.72)
NPL	152	76	881.5 (71-15675)	85.61 (3.53-9248.31)	12.09 (0.11-1251.58)	486.24 (3.5-4079.78)	1.27 (0.4-49)	0.54 (0.01-560.5)	437.42 (157.84-1055.11)	20.52 (15.62-23.46)
SAP	266	111	557.5 (36-18700)	29.3 (3.79-4050)	5.89 (0.36-197.39)	7.46 (0.58-397.69)	0.76 (0.4-56)	16.28 (0.01-5656.85)	1475.78 (889.64-2451.81)	24.76 (19.19-27.73)
SPL	174	59	1165 (145-21175)	147.12 (5.21-4351.7)	30.36 (0.43-5220)	88.89 (1.96-3716.6)	1.04 (0.5-88)	0.35 (0.01-5656.85)	616.17 (240.1-1404.61)	25.88 (11.96-32.75)
TPL	223	74	1806 (236-16219)	165.08 (11.45-1066.87)	18.9 (1.44-736.62)	34.4 (5.79-1386.77)	0.83 (0.5-47)	0.35 (0.01-5656.85)	1004.22 (376.7-1801.08)	23.33 (18.28-27.09)
UMW	201	104	1168 (195-17675)	57.4 (8.16-659)	9.97 (0.01-306.69)	9.28 (0.04-160.4)	0.62 (0.6-43)	0.35 (0.01-812.79)	960.16 (496.44-1718.2)	20.84 (17.05-23.43)
WMT	225	94	133 (22-4719)	21.84 (2.71-569.8)	0.92 (0.04-521.64)	2.97 (0.07-1682.05)	0.48 (0.3-95)	38.14 (0.01-1288.61)	669.31 (175.89-3946.33)	18.08 (11.71-29.4)
XER	196	75	344.5 (46-8000)	54.62 (4.29-4667.41)	5.35 (0.1-1867.57)	27.8 (0.02-3286.24)	1.16 (0.4-36)	2.02 (0.01-368.11)	287.58 (100.46-1176.56)	22.76 (14.69-36.11)

271 Modeling ~~taxon genus~~-environment relationships

272 We quantified relationships between macroinvertebrate assemblages and eight
273 physiochemical variables (i.e. NTL, PTL, CL, SO₄, SUBD, RPDI, MSAT, and TPRCP) using
274 joint species distribution models (JSDMs) fitted with the Hierarchical Modeling of Species
275 Communities R package (Ovaskainen and Abrego 2020, Tikhonov et al. 2020)(Ovaskainen and
276 Abrego 2020, Tikhonov et al. 2020). JSDMs, JSDMs are a multivariate hierachal generalized
277 linear mixed model fitted with Bayesian inference. They are uniquely suited to evaluate
278 relationships between anthropogenic disturbance and biological assemblages because they are
279 multi-species models that quantify taxon-environment relationships for all members in an
280 assemblage simultaneously and account for unmeasured variables, including abiotic factors and
281 species associations, using random effects specified at the sample-level (Warton et al. 2015,
282 Ovaskainen and Abrego 2020, Deflem et al. 2021)(Warton et al. 2015, Ovaskainen and Abrego
283 2020, Deflem et al. 2021). They also allow for the inclusion of phylogenetic relatedness and
284 species traits as hierarchical terms that can improve estimates of taxon-environment
285 relationships. Indeed, sample-level random effects are meaningful for multivariate models
286 because they are not confounded by residual variation as with univariate models. More
287 specifically, this attribute allows models to account for nonindependence among residuals for
288 each site and improves estimates of the fixed effects (i.e. taxon-environment relationships)
289 (Ovaskainen et al. 2016). In addition, JSDMs also allow for the inclusion of phylogenetic
290 relatedness as a hierarchical term that can lend additional improvements to estimated taxon-
291 environment relationships.

292 We provide a detailed description of the modeling framework and application elsewhere
293 (Kopp et al. 2023)(Kopp et al. 2023). In brief, the JSDM is a multivariate hierachal generalized

294 linear mixed model fitted with Bayesian inference. We fit separate models for each ecoregion
295 using genus occurrence (i.e., presence/absence) as the response variable, environmental variables
296 as fixed effects, and site-level random effects. We also use taxonomy as a surrogate for
297 phylogenetic relatedness but excluded traits to avoid complications with incomplete data and
298 increase the number of genera included in our analysis.. In brief, we used presence/absence data
299 from 1,891 benthic macroinvertebrate assemblages surveyed as part of the probabilistic and
300 handpicked sites and focused on taxa that were collected from a single site visit, identified to
301 genus, and occurred at $\geq 10\%$ of sites within an ecoregion (Table 1). Separate models were fit for
302 each ecoregion boundaries to define regional species pools and thus assume that environmental
303 conditions are the primary factor driving genus occurrence (Chessman and Royal 2004). All
304 physiochemical variables were measured in the field during the survey, except for MSAT and
305 TPRCP, which were obtained from PRISM Climate and matched to the appropriate survey year
306 (Table 1). Physiochemical variables were used as linear fixed effects and sample-level random
307 effects were used to statistically control for unmeasured variables. Previously, we found few
308 genus-environment relationships were unimodal (Kopp et al. 2023) and therefore assumed linear
309 relationships were appropriate for this study. Our analysis focused on relatively small number of
310 environmental variables because these are commonly altered by human activities. We also use
311 taxonomy as a surrogate for phylogenetic relatedness as a hierarchical term in the model. All
312 models were fitted with the default prior distributions (Ovaskainen and Abrego
313 2020)(Ovaskainen and Abrego 2020), using three independent chains (30003,000 posterior
314 samples). Convergence was determined to be satisfactory by potential scale reduction factor <
315 1.1 for fixed effects and phylogenetic parameters.

316 Model performance was evaluated with respect to the fitted models ability to reproduce
317 the observed genus richness and composition (Wilkinson et al. 2021). Predicted, taxon-specific
318 occurrence probabilities were summed to obtain predicted richness and regressed against the
319 observed richness. Our primary motivation for using these models was to measure how
320 macroinvertebrate assemblages may change if the influence of anthropogenic disturbances were
321 removed from in stream physiochemical variables and all else remained unchanged (Figure 2).
322 Since we do not use these models to predict to new locations, model performance was primarily
323 evaluated with respect to its explanatory power, i.e. the fitted models ability to reproduce the
324 observed genus richness and composition (Wilkinson et al. 2021, Abrego and Ovaskainen 2023).
325 Predicted, taxon-specific occurrence probabilities were summed to obtain predicted richness and
326 regressed against the observed richness (Calabrese et al. 2014)(Calabrese et al. 2014). We
327 determined model acceptability based on three criteria: $R^2 \geq 0.2$, $-1.5 \leq \text{intercept} \leq 1.5$, and $0.85 \leq \text{slope} \leq 1.15$ (Linke et al. 2005)(Linke et al. 2005). Compositional We calculated model
328 performance metrics using each of 3,000 posterior samples to obtain a distribution of plausible
329 estimates of performance metrics and report the mean and 5th and 95th quantiles as measures of
330 uncertainty. In addition, we assessed compositional similarity between predicted and observed
331 assemblages was assessed using a probabilistic adaptation of Jaccard similarity (Scherrer et al.
332 2020)(Scherrer et al. 2020) to avoid introducing error associated with converting predicted
333 occurrence probabilities into binary outcomes (Calabrese et al. 2014)(Calabrese et al. 2014). -We
334 calculated model performance metries similarity for each site using each of 3,000 posterior
335 samples to obtain a distribution of plausible estimates for the performance metries the mean
336 predicted occurrence probabilities and report the mean and 5th and 95th quantiles as
337 measures across all sites.

339 Predictive power for models that use random effects can only be assessed for cases where
340 at least some sampling units were included in the calibration data (Abrego and Ovaskainen
341 2023). As part of NRSA, a subsample of sites are revisited with the intent of assessing temporal
342 variability in metrics and indices (Table 1). Since we fit the JSMDs using sample-level random
343 effects, these data were used to as a second type of validation. Specifically, we evaluated
344 whether the richness observed at a revisited site was within the posterior distribution of the fitted
345 models. Although this metric of validation may seem less restrictive compared to those used to
346 evaluate explanatory power, we expected variation between samples to be rather large because of
347 stochastic events (e.g. high or low stream flows), ecological processes (e.g. emergence and
348 dispersal), and sampling procedures (e.g. field collection and laboratory subsampling) that our
349 models were not designed to capture. Given this level of potential uncertainty, this metric was
350 intended to address whether an observation is could have come from the same process that our
351 model was intended to capture. We also compared the assemblages collected during the revisit to
352 the predicted assemblages using the probabilistic adaptation of Jaccard similarity (Scherrer et al.
353 2020) and report the mean and 5th and 95th quantiles across sites.

354

355 *Modeling Physiochemical Gradients*

356 We used random forest models to relate anthropogenic and landscape geoclimatic factors
357 to total nitrogen (NTL), total phosphorous (PTL), chloride (CL), sulfate (SO₄), and substrate
358 diameter (SUBD) and make predictions if anthropogenic disturbances were removed. The suite
359 of predictor variables were selected based on their hypothesized relationship with stream water
360 chemistry and bed particle size (Lin et al. 2021, Zak et al. 2021, Kaufmann et al. 2022, Kaushal
361 et al. 2023, Sabo et al. 2023)(Lin et al. 2021, Zak et al. 2021, Kaufmann et al. 2022b, Kaushal et

362 ~~al. 2023, Sabo et al. 2023) and bed particle size.~~ Geoclimatic factors included watershed
363 morphology (e.g., basin area, elevation, and slope) and lithological characteristics (e.g.,
364 lithological phosphate, nitrogen, ~~or and~~ sulfur, and soil erodibility). Anthropogenic factors
365 included percent agriculture, road density, and presence of mines and impoundments ([Appendix](#)
366 [S1: Table S1](#)). ~~Most predictor Predictor~~ variables were obtained or derived from the StreamCat
367 Database ([Hill et al. 2016](#)) ([Hill et al. 2016](#)), National Atmospheric Deposition Program
368 ([nadp.slh.wisc.edu](#)), and EPA's National Nutrient Inventory (Sabo et al. 2019, Lin et al. 2021,
369 [Sabo et al. 2021](#), [Sabo et al. 2023](#)).

370 ~~For each gradient, the Separate random forest models were fit for NTL, PTL, CL, SO4, or~~
371 ~~SUBD using the entire national dataset to maximize the range of the predictor variables included~~
372 ~~in each model. Prior to model fitting the dataset~~ was randomly split into training and testing
373 portions (~~75/25, 80% and 20%, respectively~~) ~~and separate random forest models were fit using~~
374 ~~the complete list of candidate variables.~~) Model fit was assessed by ~~out-the coefficient of-bag~~
375 ~~determination (R².)~~ and root mean squared error (RMSE) on training and testing portions of the
376 dataset. The relative importance of the covariates was evaluated by the change in mean squared
377 error after permutating each variable (%IncMSE). ~~The presence of strong correlations between~~
378 ~~geoclimatic and anthropogenic predictor variables could generate misleading Because variable~~
379 ~~importance rankings because a can be influenced by correlated variable may become a substitute~~
380 ~~when the other is permuted (i.e., no change in %IncMSE). We did not detect strong~~
381 ~~correlations (i.e., r > 0.7) between the anthropogenievariables, we confirmed that all variables~~
382 ~~used in the models had pairwise Pearson's correlation coefficients < 0.7 and the variance~~
383 ~~inflation factors (VIF) were between 2.6 and geoclimatic variables. We then used partial3.9.~~
384 [\(Appendix S1: Table S3\)](#) In general, VIF > 5 indicates a potential problem with multicollinearity

385 (O'brien 2007). Partial dependance plots were used to visualize the relationship between the
386 most important anthropogenic factors and each environmental gradient.physiochemical variable.
387 Random forest analysis was performed using the quantregForest R Package (Meinshausen
388 2017)(Meinshausen 2017).

389

390 *Evaluating effects of human disturbance on physiochemical gradients*

391 We used the fitted random forest models to predict values for NTL, PTL, CL, SO4, and
392 SUBD if anthropogenic disturbance was removed by setting all anthropogenic factors to the
393 lowest value observed in our dataset (often zero, Appendix S1: Table S2) and leaving
394 geoclimatic factors unchanged. Hereafter “hindcast” refers to the removal of anthropogenic
395 disturbance from the physiochemical variables. Because modeled relationships have unexplained
396 residual variation, we expressed the differences between observed and hindcast values in units of
397 standard deviations (i.e., $SD = (\text{Observed} - \text{Hindcast}) / RMSE$) to quantify the effects of
398 anthropogenic disturbance relative to unexplained variation. Although we included the entire
399 range of predictor values, it is possible that hindcast data are not sufficiently similar to the data
400 used to train the model and thus susceptible to extrapolation (Meyer and Pebesma 2021, Yuan et
401 al. 2024). To test whether the hindcast dataset was sufficiently similar to the data used to train
402 the models, we first mean-centered and scaled all predictor variables to equivalent units (i.e.
403 standard deviations) and weighted them according to their importance in the model. We then
404 calculated the minimum Euclidean distance between each site in the hindcasted data, and each
405 site used in the training dataset using the same center and scale. The minimum value was then
406 divided by the mean Euclidean distance among all training data. Following Yuan et al. (2024),
407 we flagged hindcast predictions for sites that exceeded 0.5 as susceptible to extrapolation.

408 We evaluated the effects of anthropogenic disturbance on each physiochemical variable
409 based on deviations from the hindcast estimates using standardized anomalies (i.e. z-scores), or
410 the difference between the observed and hindcast value scaled by standard deviation. For
411 variables we modeled using random forests, we standardized the difference between observed
412 and hindcasted conditions using the RMSE of each model (Kilgour and Stanfield 2005)(Kilgour
413 and Stanfield 2005). We expected anthropogenic disturbance to elevate NTL, PTL, CL, and SO4
414 concentrations and increase or decrease SUBD. Accordingly, we identified sites with observed
415 concentrations greater than 2SD from the hindcast values for NTL, PTL, CL, and SO4 and either
416 2SD higher or lower for SUBD as disturbed. For climate variables, we obtained baseline
417 historical climatic values (MSAT and TPRCP) as the mean summer air temperature for 1900-
418 1950 from PRISM climate data (<https://prism.oregonstate.edu/historical/>). We then standardized
419 the difference between the present-day values (i.e. 2018 or 2019) and the baseline using the
420 standard deviation of the 50yr dataset. Importantly, using standardized anomalies rather than
421 absolute differences has advantages because it accounts for unexplained variation in the model or
422 natural variability among sites (Kilgour and Stanfield 2005). Furthermore, because the values are
423 in units of standard deviations, thresholds to evaluate whether the magnitude of difference is
424 sufficient to support an effect of anthropogenic disturbance can be rather intuitive. Specifically,
425 we expected anthropogenic disturbance to elevate NTL, PTL, CL, and SO4 concentrations and
426 increase or decrease SUBD, MSAT, and TPRCP. Accordingly, we identified sites with observed
427 concentrations >2SD from the hindcast values for NTL, PTL, CL, and SO4 and > |2SD| for
428 SUBD, MSAT, and TPRCP as having evidence of anthropogenic disturbance.
429 We also evaluated the effects of anthropogenic disturbance on climate variables and the
430 riparian disturbance index. We obtained historical climatic values (MSAT and TPRCP) from

431 PRISM 1940–1950 climate data (<https://prism.oregonstate.edu/historical/>). We selected 1940–
432 1950 climate because it predates the peak in atmospheric radiocarbon from nuclear fallout—an
433 apparent geologic marker of the Anthropocene (Lewis and Maslin 2015). For each location we
434 calculated the 10 yr mean and standard deviation and identified sites that could be disturbed
435 when differences between present day and historical values exceeded 2SD (either higher or
436 lower). Since RPDI is a direct measure of anthropogenic disturbance, we initially set this
437 variable to zero for all locations but found nearly all the sites have RPDI > 0. that this value may
438 be too strict because of the number and diversity of factors included in the index and did not
439 provide much insight into potential regional variation in disturbance. Instead, we used 0.33 as a
440 threshold which means that on average only is interpreted as one type of human disturbance
441 should be observed within 10 m of the streambanks at no more than one third of the 22
442 riparian plots sampled, on average (Kaufmann et al. 2022, USEPA 2023)(Kaufmann et al. 2022b,
443 USEPA 2023). For each environmental gradient we calculated the proportion of sites in each
444 region that Although this does mean that not all sites were completely free of riparian
445 disturbance, it is consistent with other studies that have used this index to evaluate human
446 disturbance in the riparian area (Kaufmann et al. 2022a, Kaufmann et al. 2022b, USEPA 2023).

447 For each physiochemical variable we estimated the total percent of streams that were
448 potentially affected by anthropogenic disturbance. (i.e. standardized anomaly > 2SD) using only
449 the probabilistic samples and their weights reflective of the entire population of streams and
450 rivers assigned to them by NRSA (USEPA 2023). Estimates for each ecoregion were generated
451 using the cat_analysis() function from the spsurvey R package (Dumelle et al. 2023). In
452 addition, we tallied the number of physiochemical variables that were potentially affected at each

453 site to elucidate instances where human disturbance affects multiple environmental variables
454 simultaneously.

455

456 *Evaluating effects of human disturbance on macroinvertebrate assemblages*

457 We evaluated the effects of removing anthropogenic disturbance from the
458 physiochemical environment on macroinvertebrate assemblages using the fitted JSMDs. For the
459 sites ~~with disturbed environmental variables, that had evidence of human disturbance (i.e.~~
460 ~~standardized anomaly > 2SD)~~, we ~~substitued~~substituted the hindcasted value (either predicted
461 from random forest model, ~~1940~~1900-1950 ~~elimate average~~averages for MSAT and TPRCP or
462 0.33 for RPDI) ~~for the present day value in the dataset~~ and used these data to predict
463 macroinvertebrate assemblages that could occur if disturbance was ~~absent~~removed or reduced
464 (~~Chessman and Royal 2004~~)(Chessman and Royal 2004). We then compared these hindcasted
465 assemblages to assemblages predicted under present-day conditions. To evaluate the relative ~~and~~
466 ~~additive~~ effects of ~~removing anthropogenic disturbance from the~~hindcasting each
467 physiochemical ~~environment variable~~, we created ~~54~~ scenarios changing either NTL and PTL
468 (Nutrient Scenario), CL and SO4 (Salinity Scenario), RPDI and SUBD (Habitat Scenario),
469 MSAT and TPRCP (Climate Scenario) and leaving the others at their observed ~~value or values~~. In
470 ~~addition, we predicted the macroinvertebrate assemblage after~~ changing all ~~physiochemical~~
471 variables to their hindcast value (~~Pristine~~Hindcast Scenario). For each scenario we compared
472 predicted genus richness from present-day conditions to hindcast genus richness.

473 JSMD predictions are three-dimensional data arrays that contain 3,000 posterior samples
474 of site-specific occurrence probabilities for each genus (i.e., site x genera x posterior samples).
475 For each posterior sample we summed predicted occurrence probabilities for all ~~taxa~~genera to

476 obtain 3,000 plausible estimates of hindcast genus richness for each site. We then compared the
477 mean present-day genus richness (i.e., predicted from JSMD using present-day values for the
478 physiochemical variables) to the hindcast posterior distribution ~~to evaluate.~~ We evaluated
479 whether present-day richness was below the 10th or 25th quantiles, indicating a reduction in
480 present-day genus richness relative to hindcast predictions, or above the 75th or 90th quantiles
481 (*i.e. hereafter termed quantiles, indicating an increase in genus richness relative to hindcasted*
482 (Figure 2)). Hereafter, present-day values that are located in the extreme ends of the distribution
483 are described as having either >0.75 support or >0.90 support for a difference from hindcasted
484 predictions, respectively because either 75% or 90% of the hindcast posterior distribution is
485 greater or less than the present day mean richness). We assessed the effects of removing
486 anthropogenic disturbance from each of the 4 combinations of environmental variables (*i.e.*
487 Nutrient Scenario, Salinity Scenario, Habitat Scenario and Climate Scenario) and the effects of
488 removing anthropogenic disturbance from all environmental variables simultaneously (Pristine
489 Scenario) with >0.75 and >0.90 support ~~for differences in genus richness as evidence for an effect~~
490 of disturbance on physiochemical variables.

491 Because assemblage composition could be affected by anthropogenic disturbance without
492 an subsequent change in richness (Van Sickle 2008), we also assessed compositional differences
493 between present-day and hindcast assemblages. First, we determined whether the present-day
494 mean occurrence probability of a taxon was higher or lower (>90% support) than the hindcast
495 occurrence probability for each site. Taxa that had higher occurrence probability were increasers
496 while taxa with lower occurrence probability were decreasers. Assessing compositional
497 dissimilarity between two assemblages requires binary values to indicate whether a taxon is
498 present or absent in each assemblage. We used the increaser/decreaser assignments to create two

499 community matrices, one representing present-day assemblage and the other representing the
500 hindcast assemblage. Increases were assigned a value of 1 in the present-day matrix and 0 in the
501 hindcast matrix. Conversely, decreases were assigned 0 in the present-day matrix and 1 in the
502 hindcast matrix. Taxa with <0.90 support (i.e. mean present-day occurrence probability was
503 within the 10th and 90th quantile of the hindcast distribution) were considered not to be affected
504 by anthropogenic disturbance and were assigned 1 in both matrices. We used Jaccard similarity
505 index to compare the two matrices (i.e. site x taxa) and identified compositional differences by
506 Jaccard similarity <0.9.

507 Identifying increase and decrease taxa can also elucidate differences among major
508 taxonomic groups in their response to anthropogenic disturbance and help identify which major
509 taxonomic groups are contributing most to assemblage level patterns. Accordingly, for each
510 genus we calculated the proportion of sites in which they either Identifying genera having site-
511 specific occurrence probabilities that differ between present-day and hindcast conditions helps
512 interpret assemblage-level changes in response to anthropogenic disturbance. We identified
513 genera that had higher occurrence probability under present day conditions as “increasers” and
514 genera with lower occurrence probability as “decreasers”. Differences were determined with
515 >90% support. For each genus we calculated the proportion of sites where they increased or
516 decreased. If a genus was identified as an increase at a large proportion of sites it could be
517 considered tolerant become more prevalent in response to anthropogenic disturbance.
518 Alternatively, if a genus was identified as a decrease at a large proportion of sites, it could be
519 considered sensitive become less prevalent in response to disturbance. For
520 each ecoregion, we calculated the proportion of sites wherein each genus increased or decreased
521 and then calculated report the mean proportion increasing and decreasing by for major taxonomic

522 groups (i.e.g. insects and non-insects).insect genera) to further understand taxon-specific trends
523 in the context of biodiversity loss (Jähnig et al. 2021, Rumschlag et al. 2023).

524 Identifying increaser and decreaser genera could also enhance our ability to measure the
525 effects of removing anthropogenic disturbance because assemblage composition could be
526 affected by anthropogenic disturbance without a subsequent change in richness (Van Sickle
527 2008). To assess compositional differences between present-day and hindcast assemblages we
528 used increaser/decreaser assignments to create two community matrices, one representing
529 present-day assemblage and the other representing hindcast assemblage. Increasers were
530 assigned a value of 1 in the present-day matrix and 0 in the hindcast matrix because they had a
531 higher occurrence probability under present-day conditions. Conversely, decreasers were
532 assigned 0 in the present-day matrix and 1 in the hindcast matrix because they had a significantly
533 higher occurrence probability under hindcasted conditions. For genera that did not have
534 sufficient support (< 0.90) for changing occurrence probabilities were considered not to be
535 affected by anthropogenic disturbance and were assigned 1 in both matrices. We compared the
536 two matrices (i.e. site x genera) using Jaccard similarity index and identified sites with a
537 similarity score of < 0.9 as changing compositionally.

538 We assessed the consequences of removing anthropogenic disturbance from in-stream
539 physiochemical environment for macroinvertebrate assemblages by identifying sites within each
540 region that had evidence for either a change in richness (>0.75 support) or composition (Jaccard
541 Similarity < 0.9) from present-day conditions. Then, using only the probabilistic samples that
542 were not flagged for extrapolation and had complete data (n = 1748) and their weights, estimated
543 the total percent of streams within each ecoregion that were potentially affected by
544 anthropogenic disturbance. Estimates for each ecoregion were generated using the cat_analysis()

545 function from the spsurvey R package (Dumelle et al. 2023). In addition, we report results with
546 different levels of support (> 0.75 and > 0.90) for richness change and separate compositional
547 change to convey uncertainty and methodological differences.

548

549 Results:

550 Random Forest Modeling

551 _____

552 Results:

553 Random Forest Modeling

554 The random forest models explained approximately 48 to 77% of the variation in the
555 training data (n=1502) and 51 to 78% of the variation in the test/validation data for(n=375). The
556 model of substrate diameter and ≥59% of the variation in (SUBD) had the highest RMSE in both
557 training and testing datasets compared to models of total nitrogen, (NTL), total phosphorous,
558 (PTL), chloride, (CL), and sulfate (SO₄) (Table 2). The Variable importance for each
559 physiochemical model, measured by the percent change in MSE after permutation, revealed that
560 runoff was among the most important geoclimatic variables for each model and that at least one
561 anthropogenic factor agricultural landcover in the watershed or riparian area ranked among the
562 most important anthropogenic variables (Appendix S1: Table S2). For example, percentPercent
563 agricultural landcover in the watershed was the most important variable predicting NTL and PTL
564 and the, the second most important for predicting SO₄ and third most important for predicting
565 SO₄. RoadCL. Other anthropogenic variables were also ranked relatively high in importance.
566 For example, road density in the watershed was the second most important variable predicting

567 CL and fifth most important variable for SO₄, while percent natural vegetation cover in the
568 riparian area was the third most important variable predicting SUBD.

569

570 ~~Table 2: Random Forest model performance metrics for testing and out of bag training datasets.~~
571 ~~NTL = Total Nitrogen, PTL = Total Phosphorus, CL = Chloride, SO₄ = Sulfate, SUBD =~~
572 ~~Substrate Diameter. RMSE = Root mean squared error of random forest models fitted with ln(x)~~
573 ~~+ 1) (NTL, PTL, and CL) or ln(x) (SO₄) or Log10(SUBD) transformations.~~

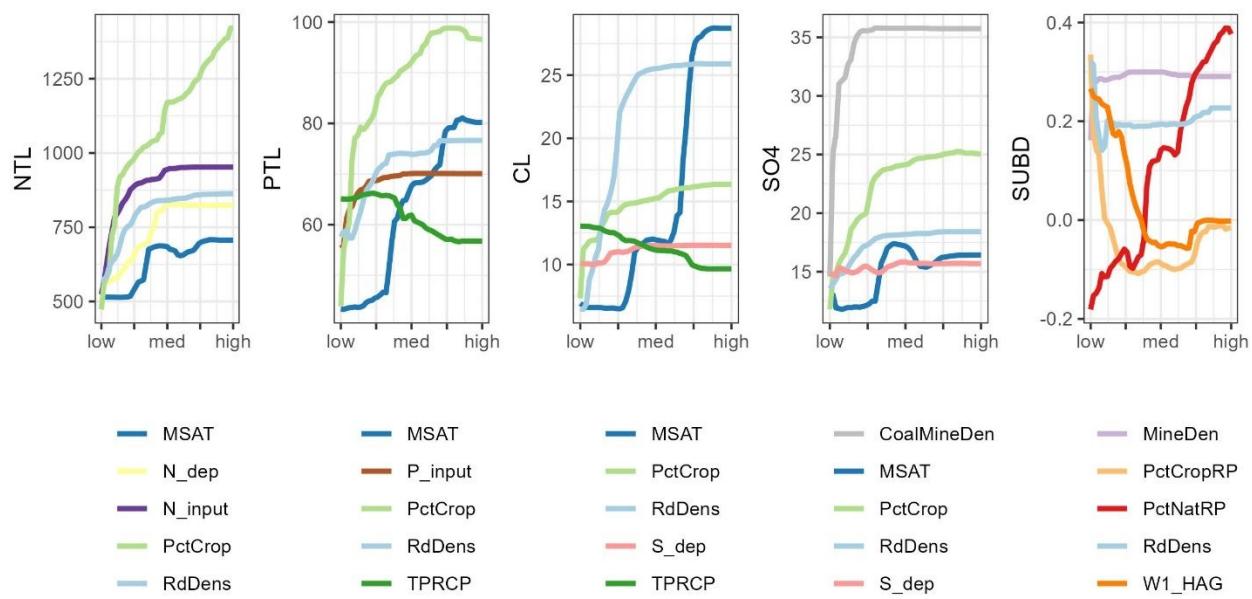
574

Variable	R ² _{train}	RMSE _{train}	R ² _{test}	RMSE _{test}
NTL (ug/L)	0.72	0.62	0.72	0.61
PTL (ug/L)	0.61	0.77	0.59	0.76
CL (mg/L)	0.70	0.77	0.71	0.82
SO ₄ (mg/L)	0.77	0.96	0.75	1.02
SUBD (mm)	0.47	0.98	0.48	1.00

575

576 We visualized the effects of the anthropogenic factors using partial dependence plots
577 (Figure 23). In general, the anthropogenic factors were associated with increased nutrients and
578 salinity and decreased substrate diameter. For example, percent crops and anthropogenic nutrient
579 inputs (i.e., of nitrogen and phosphorus)phosphorous were positively associated with total
580 nitrogenNTL and total phosphorous concentrationsPTL. Road density and mean summer air
581 temperature had a positive association with chlorideCL while densitytotal annual precipitation
582 had a weakly negative relationship. Density of coal mines in the watershed was positively
583 associatedhad a relatively strong positive effect on SO₄ and SUBD had a strong positive
584 relationship with sulfate. Substrate diameter increased with percent natural vegetation cover and
585 strong negative relationship with agricultural activities in the riparian areaand decreased with
586 percent agriculture and anthropogenic disturbance.

587



588

589 **Figure 2: Partial dependence plots showing the effects of the anthropogenic variables on each**
 590 **environmental gradient. For visualization, each anthropogenic factor was rescaled between 0-1**
 591 **and labeled low, medium, and high. CoalMineDen = coal mine density, MineDen = gravel mine**
 592 **density, MSAT = Mean summer air temperature, N_dep = atmospheric nitrogen deposition,**
 593 **N_input = Anthropogenic nitrogen inputs, P_input = Anthropogenic phosphorous inputs,**
 594 **PetCrop = Percent crop in the watershed, PetCropRP = percent crop in the riparian area,**
 595 **PetNatRP = percent natural vegetation in riparian area, RdDen = Road density, S_dep =**
 596 **atmospheric sulfur deposition, TPRCP = total precipitation, and W1_HAG = agricultural**
 597 **disturbance adjacent to stream reach. NTL = Total Nitrogen (ug/L), PTL = Total Phosphorus**
 598 **(ug/L), CL = Chloride (mg/L), SO4 = Sulfate (mg/L), SUBD = Substrate Diameter Log10(mm).**

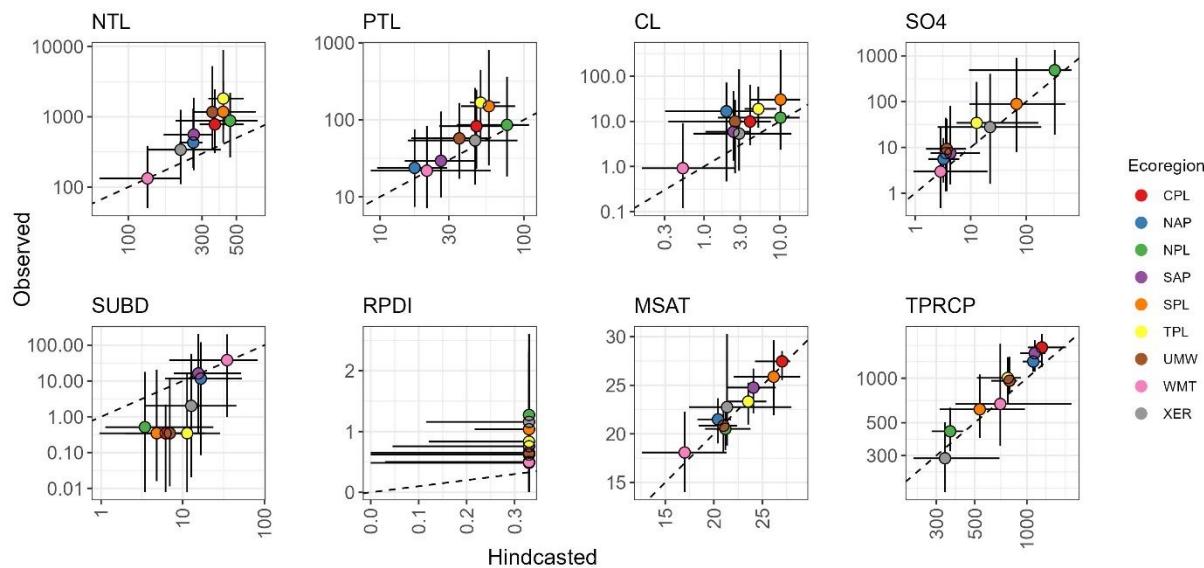
599

600

601 *Disturbance effects on physiochemical conditions*

602 Removing anthropogenic disturbance from the physiochemical environment generally
 603 decreased nutrient concentrations, salinity, and riparian disturbance and increased substrate
 604 diameter (Figure 3). Change in total nitrogen and phosphorus concentrations were most evident
 605 in Temperate Plains and Southern Plains and change in chloride was most evident in the
 606 Northern Appalachians. Substrate diameter was noticeably coarser after hindcasting for all
 607 regions except Erie, Southern Appalachians, and Northern Appalachians. Observed and

608 hindcasted values for sulfate, mean summer air temperature and total precipitation were
 609 generally similar for all ecoregions.

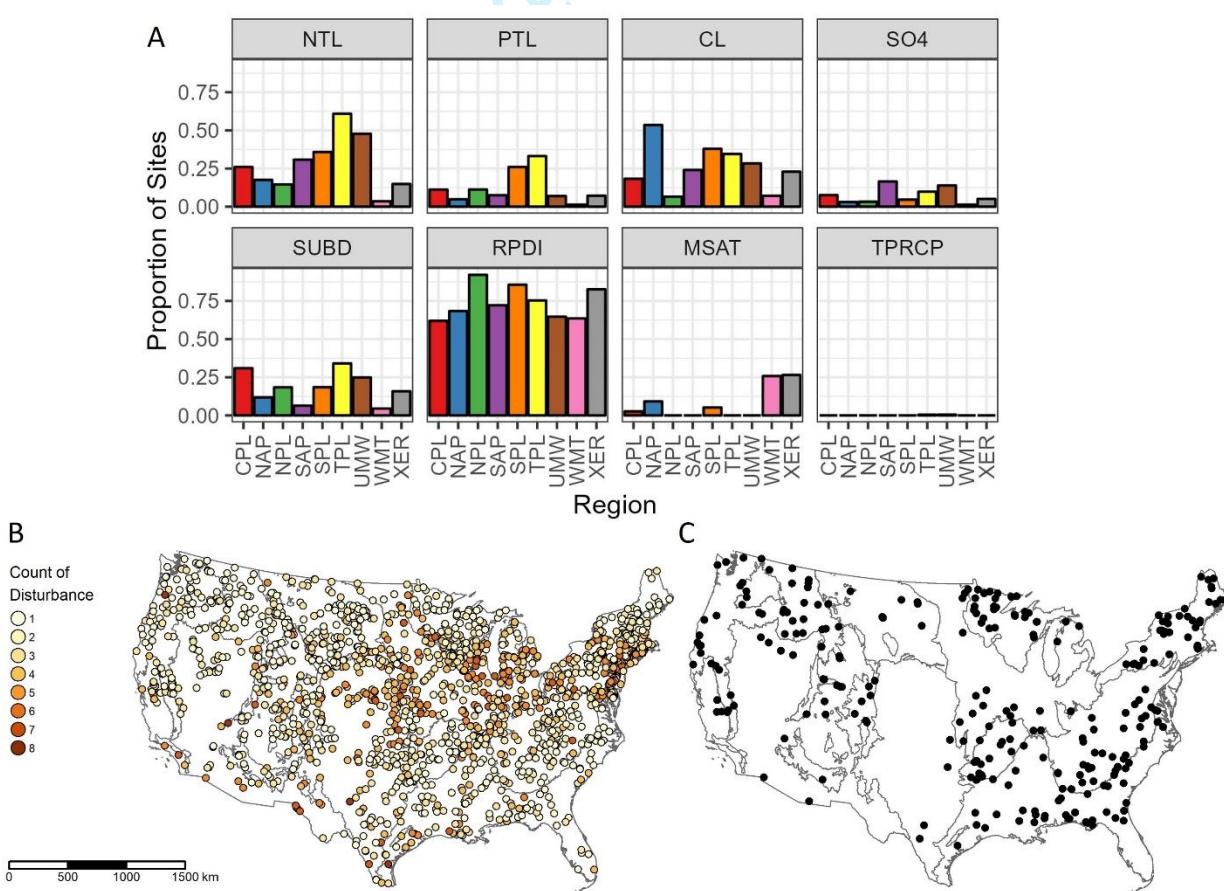


610
 611 **Figure 3:** Ecoregional values of observed versus hindcasted values for each environmental
 612 gradient. Points are regional means and vertical and horizontal bars represent the 10th and 90th
 613 quantiles of observed values or hindcasted values within each region, respectively. The bars
 614 demarcate the typical range of values for each environmental variable within a region under each
 615 scenario. The dashed line is the 1:1 relationship, such that points above the line indicate that, on
 616 average, present-day values for the environmental variables within an ecoregion are greater than
 617 the average hindcasted value. NTL = Total Nitrogen (ug/L), PTL = Total Phosphorus (ug/L), CL
 618 = Chloride (mg/L), SO4 = Sulfate (mg/L), SUBD = Substrate Diameter (mm), RPDI = Riparian
 619 Disturbance Index, MSAT = mean summer air temperature (°C) and TPRCP = total precipitation
 620 (mm). CPL = Coastal Plains, NAP = Northern Appalachians, NPL = Northern Plains, SAP =
 621 Southern Appalachians, SPL = Southern Plains, TPL = Temperate Plains, UMW = Upper
 622 Midwest; WMT = Western Mountains; XER = Xeric. See Table S3 for observed and hindcasted
 623 mean values and quantiles for each region.

624

625 Anthropogenic disturbance affected physiochemical gradients differently depending on
 626 the ecoregion (Figure 4a). For example, the difference between present-day and hindcasted total
 627 nitrogen concentrations exceeded 2 standard deviations in 60% of the sites in Temperate Plains
 628 and 47% of sites in Upper Midwest. Similarly, 53% of the sites in the Northern Appalachians
 629 were found to have potentially elevated CL concentrations and >30% of sites in Coastal Plains
 630 and Temperate Plains had finer substrates. Most sites exceeded the 0.33 threshold for RPDI

631 based on the observed and hindcasted values while relatively fewer sites exceeded thresholds for
 632 sulfate concentrations, mean summer air temperature, or total precipitation. Among the
 633 ecoregions, Southern Appalachians had the largest proportion of sites with elevated sulfate
 634 concentrations relative to hindcasted values and several sites in Xeric and Western Mountains
 635 had historic MSAT higher than present day. Anthropogenic disturbance could alter multiple
 636 physiochemical variables simultaneously (Figure 4b). For example, >50% of sites in Northern
 637 Appalachians, Southern Plains, Temperate Plains, and Upper Midwest had >2 physiochemical
 638 gradients change as a result of removing anthropogenic disturbance. Conversely, Western
 639 Mountains and Southern Appalachians had the greatest proportion of sites (>20%) without
 640 disturbance.



641

642 ~~Figure 4: A) Proportion of sites where hindcasted abiotic conditions were >2 standard deviations~~
643 ~~from the observed value for Total Nitrogen (NTL), Total Phosphorus (PTL), Chloride (CL),~~
644 ~~Sulfate (SO₄), Substrate Diameter (SUBD), mean summer air temperature (MSAT), and Total~~
645 ~~Precipitation (TPRCP) and > 0.33 for Riparian Disturbance Index (RPDI). B) The number of~~
646 ~~environmental variables where differences between hindcasted and observed abiotic conditions~~
647 ~~exceeded the threshold. C) Locations where differences between hindcasted and observed abiotic~~
648 ~~conditions did not exceed the threshold for any variable. CPL = Coastal Plains, NAP =~~
649 ~~Northern Appalachians, NPI = Northern Plains, SAP = Southern Appalachians, SPL = Southern~~
650 ~~Plains, TPL = Temperate Plains, UMW = Upper Midwest, WMT = Western Mountains, XER =~~
651 ~~Xeric.~~

652

653 *JSDM assemblage level performance* ——

654 ~~Depending on the ecoregion, we~~
655 ~~We fitted JSDMs for each region using~~
656 ~~presence/absence data for 59 – 127 genera using 127 – 255 from 152 – 266 sites surveyed during~~
657 ~~the 2018 – 2019 NRSA cycle (Table 1). The total number of sites used to fit JSDMs were higher~~
658 ~~(1,891) than the sites we ultimately assessed (1,824) because we included all available~~
659 ~~information to maximize the number of occurrences for each genus probabilistic and handpicked~~
660 ~~sites (USEPA 2023)(USEPA 2023)~~. We found that summing predicted occurrence probabilities
661 ~~for all taxagenera modeled overestimated could potentially overestimate~~ observed genus richness
662 ~~(i.e. observed vs predicted richness intercepts < -1.5).~~ Upon further inspection we found that this
663 ~~overprediction was due to many taxagenera having exceptionally low predicted occurrence~~
664 ~~probabilities potentially owing to the large spatial extent of our ecoregions and but nonetheless~~
665 ~~present in~~ the regional ~~taxaspecies~~ pool. Thus, we ~~used a threshold to~~ excluded ~~taxagenera with~~
666 ~~predicted occurrence probabilities < 0.05 or < 0.10 prior to summation to correct this bias (Table~~
667 ~~3).~~

667 ~~Because site level random effects accounted for unmeasured factors in the models, we did~~
668 ~~not expect our models to have strong predictive power when projected to new locations (Abrego~~
669 ~~and Ovaskainen 2023, Kopp et al. 2023).~~ However, we could assess whether values obtained at
670 ~~revisited sites were within the predicted posterior distribution of the fitted models as a secondary~~

671 form of validation. Although the number of revisited sites available for model testing was low
672 (Table 1), we found that >90% of the genus richness values observed during revisit sampling
673 were within the posterior distribution of the models. Importantly, this lends plausibility that the
674 posterior distribution reflects processes characterizing macroinvertebrate assemblages (Table 3
675 and Appendix S1: Figure S2).

676 The probabilistic adaptation of Jaccard similarity revealed that the predicted and
677 observed assemblages were on average 48–56% similar (Table 3).>50 similar (Table 3). Sites
678 located the WMT tended to have the most similar assemblages (mean = 0.6, 5th quantile = 0.36,
679 95th quantile= 0.75) while sites in the XER tended to have the least similarity (mean = 0.23, 5th
680 quantile = 0.23, 95th quantile= 0.71). Although potentially vulnerable to sample size constraints,
681 mean similarity between model predictions and assemblages observed at revisited sites were also
682 relatively high, with mean values ranging between 0.35 and 0.68 (Table 3).

683

684 *Disturbance effects on physiochemical conditions*

685 We tested whether using the random forest models to predict physiochemical conditions
686 after removing the effects of anthropogenic activities were susceptible to extrapolation. We
687 found predictions at 77 of the sites (~4%) were potentially susceptible to extrapolation. This
688 indicates that removing disturbance would cause some sites to be sufficiently different from the
689 data used to train the models and suggest that these sites do not have a natural analog. Indeed,
690 most of the sites flagged for potential extrapolation were located in the Coastal Plains Ecoregions
691 (18%) and were associated with models used to predict NTL concentrations in the lower
692 Mississippi and SUBD along the southern coast (Appendix S1: Figure S1). To avoid the

693 potential for bias associated with extrapolation, we designated them as “Not Assessed” in our
 694 analysis.

695 For sites that were not potentially susceptible to extrapolation, removing anthropogenic
 696 disturbance from the physiochemical environment generally decreased nutrient concentrations,
 697 salinity, and riparian disturbance and increased substrate diameter (Figure 4). Change in NTL
 698 and PTL concentrations were most evident in Temperate Plains and Southern Plains and change
 699 in CL was most evident in the Northern Appalachians. Substrate diameter was noticeably coarser
 700 after hindcasting for all regions except Xeric, Southern Appalachians, and Northern
 701 Appalachians. Table 3. Regression coefficients between predicted and observed richness and
 702 compositional similarity. Occurrence probabilities thresholds (Pr) were used to exclude taxa that
 703 with low predicted occurrence probabilities. We considered $R^2 \geq 0.2$, $-1.5 \leq \text{intercept} \leq 1.5$, and
 704 $0.85 \leq \text{slope} \leq 1.15$ are indicative of adequate model performance (Linke et al. 2005). Values in
 705 parentheses are the 5th and 95th percentile of estimates from the 3000 posterior samples.

	Pr	Intercept	Slope	R^2	Mean Jaccard Similarity
CPL	0.05	0.37 (-0.37–1.02)	1.00 (0.96–1.05)	0.85 (0.82–0.87)	0.49 (0.24–0.65)
NAP	0.05	-0.45 (-2.02–1.02)	1.03 (0.99–1.08)	0.74 (0.71–0.78)	0.51 (0.29–0.63)
NPL	0.05	-0.07 (-1.21–1.04)	1.02 (0.96–1.09)	0.76 (0.71–0.81)	0.56 (0.33–0.73)
SAP	0.05	0.40 (-0.66–1.41)	1.00 (0.97–1.04)	0.85 (0.82–0.87)	0.5 (0.25–0.65)
SPL	0.10	0.48 (-0.45–1.29)	1.02 (0.97–1.09)	0.79 (0.75–0.83)	0.48 (0.2–0.68)
TPL	0.05	0.06 (-0.91–0.94)	1.01 (0.97–1.06)	0.82 (0.79–0.84)	0.52 (0.28–0.72)
UMW	0.05	-0.74 (-2.52–0.99)	1.04 (0.98–1.10)	0.73 (0.66–0.78)	0.52 (0.33–0.66)
WMT	0.10	0.24 (-1.24–1.59)	1.04 (0.99–1.10)	0.74 (0.70–0.78)	0.55 (0.32–0.72)
XER	0.05	-0.12 (-1.11–0.74)	1.03 (0.98–1.08)	0.83 (0.80–0.86)	0.48 (0.18–0.66)

706
 707
 708 Observed and hindcasted values for sulfate, and mean summer air temperature were
 709 generally similar for all ecoregions.

710 Anthropogenic disturbance affected physiochemical gradients differently depending on
 711 the ecoregion (Figure 5a). We found 60.8% (95% Confidence Interval, hereafter CI = 52.8–

712 68.8%) of streams in Temperate Plains and 38.4% (CI = 32.0-44.9%) of streams Upper Midwest
713 had elevated NTL relative to hindcasted estimates. In the Northern Appalachians, 41.4% (CI =
714 34.5-48.2%) of streams were found to have potentially elevated CL concentrations. Streams with
715 excess fine substrates were most prevalent in the Temperate Plains (29.2%, CI = 21.2-37.1%)
716 and Northern Plains (21.4%. CI = 14.1-28.9%). We also found similarities among regions. RPDI
717 and TPRCP were the most prevalent disturbance among all ecoregions while disturbed SO4 and
718 MSAT values were the least prevalent. Finally, anthropogenic disturbance could alter multiple
719 physiochemical variables simultaneously (Figure 5b). In general sites with >2 physiochemical
720 variables disturbed by anthropogenic activities were located in Northern Appalachians, Southern
721 Plains, Temperate Plains, and Upper Midwest.

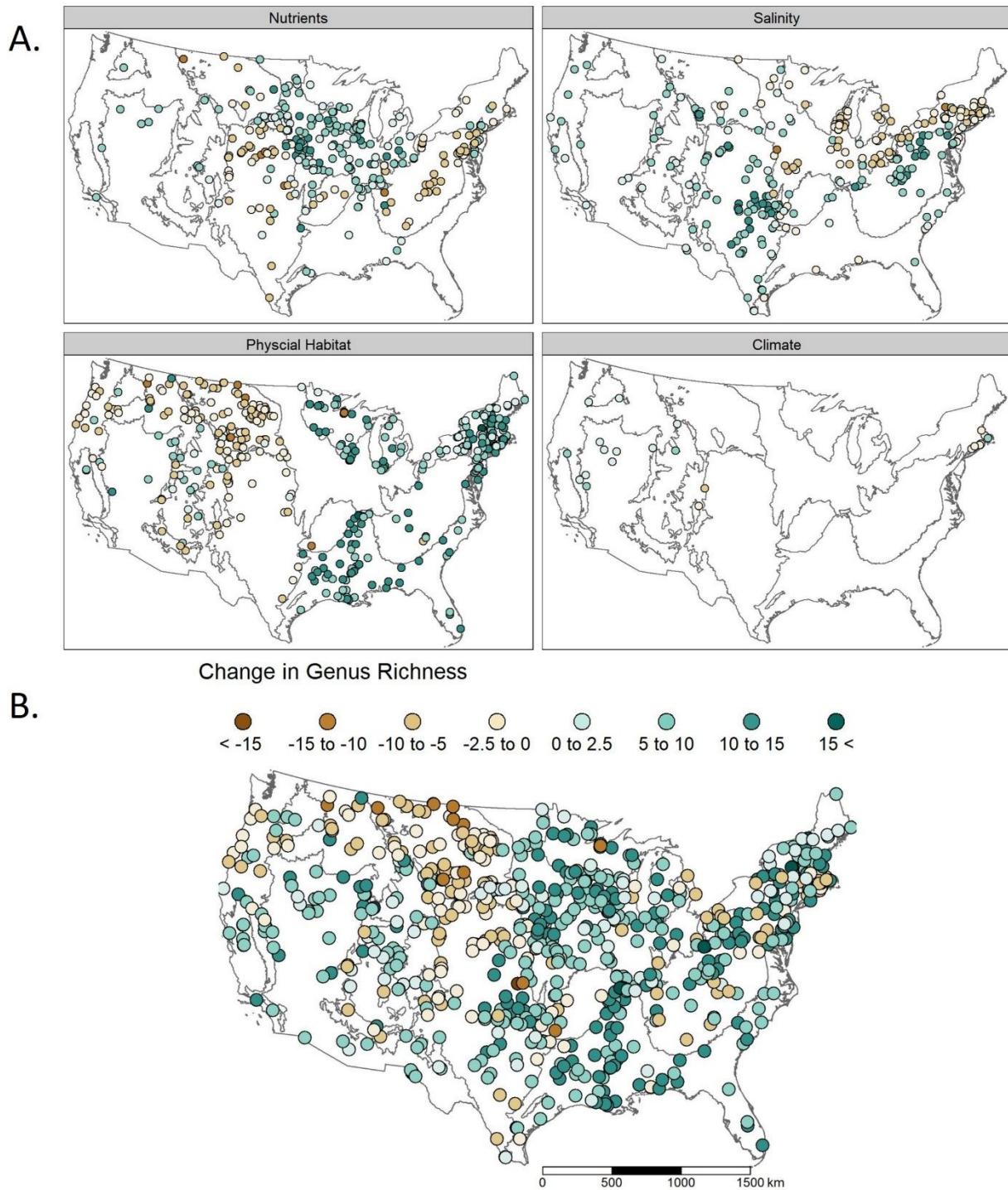
722

723 *Macroinvertebrate assemblage response to undisturbed physiochemical environment*
724 We evaluated how Macroinvertebrate models were fit for each region separately and
725 we assessed whether removing anthropogenic disturbance would produce estimates that exceed
726 the range of values used to fit each model. Hindcast predictions exceeded the range of present-
727 day conditions at 1 location in CPL, SAP and XER, 6 locations in TPL, and 12 locations in
728 WMT. In these instances, we reset the hindcasted values to be within the range (minimum) of
729 data used to fit the mode to avoid potential extrapolation.

730 Hindcast predictions for each category of physiochemical variables generated different
731 outcomes for macroinvertebrate assemblages would respond to removing anthropogenic
732 disturbance from the abiotic variables using 5 scenarios. The effects of each scenario varied
733 among regions and could increase or decrease genus richness (Figure 5a). (Figure 6a). For
734 example, the nutrient scenario (removing anthropogenic disturbance from NTL and PTL), tended

735 ~~to increase concentrations, produced macroinvertebrate assemblages with higher genus richness~~
736 in ~~the~~ Temperate Plains ~~but decrease and~~ lower richness at some sites in the Southern Plains and
737 Southern Appalachians ecoregions. ~~The salinity scenario (Figure 6a). On the other hand,~~
738 ~~removing disturbance from CL and SO₄~~ tended to increase genus richness in the Southern Plains
739 but ~~decrease decreased~~ richness at sites in the Northern Appalachians. ~~The Hindcasted physical~~
740 habitat ~~scenario, tended to increase richness in the Coastal Plains, Upper Midwest and Northern~~
741 ~~Appalachians and decrease richness in the Northern Plains and Western Mountains. Finally, the~~
742 ~~climate scenario had a relatively small effect on variables, mostly affected macroinvertebrate~~
743 ~~assemblages in the Northern Plains by decreasing richness but also increased~~ genus richness ~~in~~
744 ~~Coastal Plains, and Northern Appalachians and Upper Midwest (Figure 6a). TPRCP generally~~
745 ~~contributed to changes in macroinvertebrate assemblages~~ because relatively few sites had
746 ~~differences between observed and 1940–1950 averages for MSAT and TPRCP temperature~~
747 ~~anomalies exceed > 2 standard deviations. The pristine scenario changed all abiotic gradients to~~
748 ~~reflect the removal of disturbance and from 1900–1950 mean. In the Southern Appalachians~~
749 ~~genus richness decreased from present day conditions while in the Northern Plains genus~~
750 ~~richness generally increased from present day conditions. Removing hindcasting all~~
751 ~~physiochemical variables~~ typically increased genus richness (Figure 5b). Collectively these
752 results suggest that human alteration of nutrient and salinity concentrations and habitat typically
753 reduces macroinvertebrate assemblage richness in all regions except for the Northern Plains and
754 Southern Appalachians (Figure 6b).

755



756

757 **Figure 5:** Difference in macroinvertebrate genus richness predicted from present-day and after
 758 removing anthropogenic disturbance from each category of environmental variable (A) and all
 759 variables simultaneously (B). The points in each panel are sites that had a change in genus
 760 richness with >75% support after removing disturbance.

761

762 —————Identifying genera that had higher present-day occurrence probabilities (increasers), or
763 higher hindcast occurrence probabilities (decreasers) ~~elucidates how major taxonomic groups~~
764 ~~may respond to disturbance (Table 4). For example, if disturbance was removed from the~~
765 ~~physiochemical variables, our results indicate that occurrence probabilities of Ephemeroptera,~~
766 ~~Plecoptera and Trichoptera (EPT) would on average increase at 26% of the sites in CPL.~~
767 ~~Alternatively, members of Mollusca in NPL tend to increase at 33% of the sites. More generally,~~
768 ~~insects in CPL, NAP, SAP, TPL, UMW and XER tended to be decreasers while non-insects~~
769 ~~tended to be increasers at more sites in all regions except CPL.~~

For Review Only

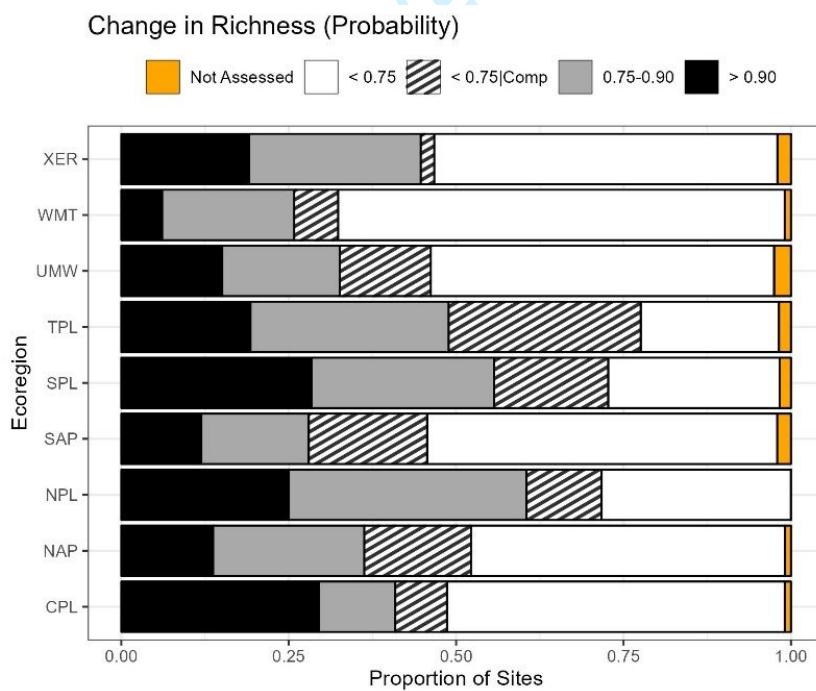
770 indicated that insects tended to be decrease~~s~~rs while non-insects tended to be increase~~s~~rs, but this was not consistent across all
771 regions (Table 4). For example, in the Southern Plains and Western Mountains, insects were almost equally likely to be increase~~s~~rs (0.1
772 and 0.6, respectively) or decrease~~s~~rs (0.09 and 0.03, respectively) and in the Northern Plains both insects and non-insects tended to be
773 increase~~s~~rs at a larger proportion of sites (0.19 and 0.22). Among insects, members of chironomidae were generally most likely to be
774 increase~~s~~rs. Among non-insects, in contrast, members of mollusca were most likely increase~~s~~rs (Table 4). Thus, higher present-day
775 genus richness compared to hindcast estimates could be due to genera that are typically considered to be tolerant of human activities.

776 Using the probabilistic sites, we aggregated our results to reflect the total population of streams and rivers in each ecoregion
777 and detected that 14.3 – 75.5% of streams have present-day assemblages that differ from assemblages expected from the hindcasted
778 physiochemical environment (Figure 7). With respect to changes in richness with support >0.75, the Northern Plains (58.3%),
779 Temperate Plains (54.8%), and Southern Plains (54.2%) the most streams with potentially altered macroinvertebrate assemblages
780 while the Western Mountains (9.2%), Upper Midwest (19.6%), and Northern Appalachians (29.7%) have the fewest. Compositional
781 change without a corresponding change in genus richness could increase the percentage of streams with altered assemblages. Evidence
782 for compositional change was most pronounced in Temperate Plains (20.7%), Southern Appalachians (20.7 %) and Upper Midwest
783 (18.8%). All regions had a percentage of streams that could not be assessed because of potential extrapolation of hindcasted
784 physiochemical variables, incomplete data or macroinvertebrate assemblages consisting of only rare genera (i.e. prevalence < 10%).
785 Among them, the Coastal Plains (12.9%) and Xeric (11.1%) had the largest percentage of streams that were not assessed, thus the total
786 number of sites assessed was less than the total number of probabilistic sites surveyed (n = 1748).

787 ~~Table 4: Mean proportion of sites in each taxonomic group that was identified as an increaser or decreaser. Increasers (+) are genera~~
 788 ~~that have a significantly higher probability of occurrence under present day conditions compared to hindcasted conditions; decreasers~~
 789 ~~(-) are genera that have a significantly lower probability of occurrence under present day conditions compared to hindcasted~~
 790 ~~conditions. The values in the table are the mean proportion of sites for genera within the major taxonomic group. CPL = Coastal~~
 791 ~~Plains, NAP = Northern Appalachians, NPL = Northern Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL =~~
 792 ~~Temperate Plains, UMW = Upper Midwest; WMT = Western Mountains; XER = xeric. EPT = Ephemeroptera, Plecoptera and~~
 793 ~~Trichoptera.~~

	CPL		NAP		NPL		SAP		SPL		TPL		UMW		WMT		XER	
	D	+	D	+	D	+	D	+	D	+	D	+	D	+	D	+	D	+
Insects	0.17	0.08	0.07	0.05	0.08	0.19	0.10	0.05	0.09	0.10	0.17	0.11	0.12	0.07	0.03	0.06	0.11	0.06
—EPT	0.26	0.00	0.08	0.04	0.12	0.14	0.12	0.04	0.09	0.03	0.24	0.04	0.16	0.01	0.05	0.05	0.13	0.01
—CHIRONOMIDAE	0.13	0.10	0.07	0.05	0.02	0.19	0.05	0.06	0.03	0.13	0.11	0.14	0.06	0.13	0.02	0.07	0.07	0.10
—Other Insects	0.17	0.06	0.07	0.07	0.06	0.24	0.11	0.05	0.21	0.03	0.20	0.11	0.10	0.02	0.03	0.02	0.13	0.01
Non-Insects	0.14	0.13	0.04	0.17	0.05	0.22	0.07	0.11	0.05	0.17	0.12	0.13	0.07	0.13	0.04	0.08	0.04	0.12
—ARTHROPODA	0.23	0.12	0.04	0.18	0.09	0.01	0.08	0.09	0.07	0.07	0.20	0.24	0.12	0.00	0.05	0.04	0.04	0.04
—MOLLUSCA	0.08	0.10	0.05	0.21	0.01	0.33	0.09	0.14	0.09	0.09	0.11	0.08	0.08	0.05	0.01	0.07	0.01	0.20
—Other Non-Insects	0.19	0.16	—	0.13	0.01	0.15	0.01	0.11	0.01	0.24	0.06	0.13	0.01	0.24	0.04	0.12	0.06	0.11

795 We identified sites that had a Jaccard similarity < 0.9 to assess compositional changes at
 796 sites that did not change in richness. Changes in composition were most pronounced in TPL,
 797 SPL, SAP, and NAP ($> 15\%$ of sites). Adding sites with evidence of compositional change to
 798 sites that had evidence of a change in genus richness revealed that removing anthropogenic
 799 genus richness or composition at 32–78% of the sites within an ecoregion (Figure 6).
 800 Macroinvertebrate assemblages would change the least in the WMT ecoregion and the most in
 801 the NPL, TPL and SPL. Considering only changes in richness revealed similar patterns but the
 802 proportion of sites identified to potentially change depended on the level of support. Considering
 803 sites with > 0.75 support for differences in richness revealed that disturbance could change the
 804 number of genera at 26–61% while the more conservative threshold (i.e. > 0.9 support) revealed
 805 that 6–30% sites might change in richness.
 806



808 Figure 6: Proportion of sites where genus richness or composition could change if anthropogenic
 809 disturbance was removed from the physiochemical environment. Shaded bars indicate that the
 810 probability of mean present day richness differs from hindcast with > 0.90 (Black) or $0.75–0.90$

811 (Gray) support. Hatched bars indicate that compositional similarity was < 0.9 but support for a
812 difference in richness were < 0.75 . White bars indicate the proportion of sites that had < 0.75
813 support for change in richness and > 0.9 compositional similarity. Yellow bars indicate the
814 proportion of sites that could not be assessed because of insufficient data.

815

816 Discussion

817 Assess the effects of anthropogenic disturbance on biological assemblages is challenging
818 because few minimally disturbed sites remain (Stoddard et al. 2006). Sites in least-disturbed are
819 often adopted as an alternative but are difficult to consistently define, vary in quality, and can be
820 spatially aggregated (Hawkins et al. 2010, McNellie et al. 2020). The model-based framework
821 we developed contributes to a suite of other approaches intended to circumvent the need for
822 reference sites (Chessman and Royal 2004, Elias et al. 2016, Yuan et al. 2024). Specifically, we
823 evaluated the effects of human disturbance on several physiochemical variables and, in turn,
824 addressed whether altered physiochemical conditions affect macroinvertebrate assemblages.
825 Using this approach, we found that anthropogenic disturbance can affect multiple
826 physiochemical variables simultaneously and that the effects on any single factor can vary
827 among ecoregions. We also found that removing or reducing disturbance could change genus
828 richness at $> 50\%$ of the streams in some ecoregions and up to 75% if compositional change is
829 considered. Collectively, our framework offers a promising alternative to evaluating the effects
830 of specific disturbances on macroinvertebrate assemblages that does not rely on reference sites.

831

832 *Hindcasting physiochemical conditions*

833 Random forest models have been used by others to predict physiochemical variables if
834 human disturbance was reduced or removed (Yuan et al. 2024). In general, using models to infer
835 physiochemical conditions eliminates the need to identify reference sites or collect excessive

836 data from undisturbed locations (Herlihy and Sifneos 2008, Soranno et al. 2011). In this study,
837 random forest models improve on earlier regression-based hindcasting approaches (Dodds and
838 Oakes 2004, Herlihy and Sifneos 2008) because they incorporate a relatively large number of
839 natural and anthropogenic predictor variables and accommodate their complex relationships with
840 the response variables (Yuan et al. 2024). Importantly, because reference sites may not be
841 representative of all sites that need to be assessed including natural and anthropogenic variables
842 enabled us to make site-specific predictions rather than relying on an estimated mean value (i.e.
843 the intercept from a multiple regression as a function of only disturbance) for an entire region
844 (Dodds and Oakes 2004, Herlihy and Sifneos 2008). Thus, our hindcasting approach provides a
845 first-order approximation for what conditions at a site could be if they are presently unknown.

846 Furthermore, random forest models permitted us to evaluate the relative importance of
847 specific anthropogenic activities and characterize potential site and regional scale differences
848 after reducing or removing them. For example, higher than expected total nitrogen in Temperate
849 Plains and Upper Midwest and the relatively high importance of percent agriculture landcover
850 our model suggests that nutrient inputs associated with agricultural activities may be elevating
851 in-stream nutrient concentrations (Lin et al. 2021, Sabo et al. 2023). Similarly, elevated chloride
852 concentrations in the Northern Appalachians and the relative high importance of road density in
853 the model, could reflect the contribution of road salts to freshwater salinization (Kaushal et al.
854 2023).

855 Certainly, inferring physiochemical conditions for different levels of human disturbance
856 depends on the quality and structure of the model. Although our models were evaluated using an
857 independent validation dataset, hindcasting required a dataset that reduced or eliminated the
858 effect of predictor variables associated with human disturbance. As consequence, the dataset

859 used for hindcasting could be dissimilar from the data used to calibrate the model and potentially
860 generate errors associated with extrapolation (Meyer and Pebesma 2021). For most regions, the
861 hindcasting dataset was sufficiently similar to the training dataset such that we could assess the
862 alteration of physiochemical variables for majority of sites. Nonetheless 18% of sites surveyed
863 the Coastal Plains were flagged as a result of testing for extrapolation. Considering that many of
864 the sites were located in the lower Mississippi River Basin, a unique and heavily modified
865 system, it is perhaps unsurprising that removing human disturbances from this system is beyond
866 the domain of our model. Testing for potential extrapolation in the context of hindcasting is
867 novel and, although the threshold to determine whether our results were prone to extrapolation
868 has been used previously (Yuan et al. 2024), more research should be devoted to optimizing and
869 interpreting this threshold. When locations or streams are potentially susceptible to extrapolation,
870 it might be more reasonable to select values that represent management goals, best professional
871 judgement or those that maximize societal benefits (Bouleau and Pont 2015).

872 To assess how human-related disturbance effects physiochemical environment, we used
873 standardized anomalies (i.e. z-scores), to account for either unexplained variation in the model
874 values or variation along a baseline. Thresholds are of central importance for the communication
875 and evaluation of human disturbances communication (Wood 2008). Here, we used a $\pm 2SD$
876 threshold for deciphering whether human disturbances alter physiochemical conditions (Kilgour
877 and Stanfield 2005). Although this threshold is intuitive, it could potentially be too strict for
878 some variables. For example, values for MSAT at many locations did not exceed the threshold in
879 our analysis, suggesting that 2018 and 2019 temperature averages are consistent with 1900-1950
880 averages, given interannual variation. However, warming of 1SD has been implicated in
881 reductions of insects in agricultural landscapes (Outhwaite et al. 2022) and there is evidence for

882 increasing summer air temperatures in the United States of approximately 0.09°C per decade
883 since 1901 (USEPA 2024a). On the other hand, to evaluate riparian disturbance, we used a
884 threshold of 0.33 which suggests that human disturbance is not completely absent for all sites
885 that were below this threshold. Indeed, nearly every site in our analysis has RPDI >0 such that
886 this threshold did not informatively differentiate regional variation in disturbance. Further, this is
887 the threshold used by others to identify least disturbed sites (Kaufmann et al. 2022b, USEPA
888 2023) and we sought to evaluate how this could affect microinvertebrate assemblages. Although
889 thresholds were necessary to identify which regions may have relatively more disturbed
890 physiochemical conditions that others, future applications may select different thresholds, report
891 continuous scores, or vary the amount of disturbance to meet specific applications (Yuan et al.
892 2024).

893

894 *Application for biological assessment without reference sites*

895 ~~Modeling has improved our ability to quantify the effects of anthropogenic disturbance~~
896 ~~on biological assemblages (Hawkins et al. 2000, Hawkins 2006, Bailey et al. 2014). However,~~
897 ~~these approaches have traditionally relied on collecting reference assemblages from sites that are~~
898 ~~representative of a least disturbed state. Because streams and rivers are among the most degraded~~
899 ~~ecosystems (Dudgeon et al. 2006, Vörösmarty et al. 2010), sites in least disturbed condition can~~
900 ~~be spatially aggregated or vary in quality among ecoregions (Herlihy et al. 2008, McNellie et al.~~
901 ~~2020). This makes it challenging to elucidate the full extent of anthropogenic disturbance on~~
902 ~~biological assemblages. Two-stage modeling approaches that combines estimates of abiotic~~
903 ~~conditions without human disturbance taxon-environment relationships could improve our~~
904 ~~understanding of how disturbance affects macroinvertebrate assemblages (Yuan et al. 2024). Our~~

905 approach builds on previous approaches by using a JSMD framework. We found that
906 anthropogenic disturbance can simultaneously affect multiple physiochemical gradients and that
907 removing disturbance could change genus richness at >50% of the sites and up to 75% if
908 compositional change is considered. Typically, sites increased in richness under hindcast
909 conditions, indicating that genus richness is presently lower than it would be without human
910 disturbance. Because JSMD also generate occurrence probabilities for each taxon, we assessed
911 taxon-specific differences between hindcasted and present-day conditions predictions for each
912 site. In general, we found that insect taxa typically have lower predicted probabilities for more
913 sites under present-day conditions compared to non-insects. This could support the notion that
914 insect taxa tend to decrease in the presence of anthropogenic disturbance while non-insect taxa
915 increase (Rumschlag et al. 2023).

916

917 *Hindcasting physiochemical conditions*

918 Anthropogenic disturbance affects macroinvertebrate assemblages by modifying the in
919 stream physiochemical environment. Using hindcast physiochemical variables to predict the
920 assemblage expected to occur if anthropogenic disturbance was removed allowed us to infer the
921 consequences of altering a relatively small number of physiochemical variables. Central to our
922 application is that multivariate models can implement meaningful site-level random effects that
923 statistically control for unmeasured environmental variables and potential biotic interactions
924 (Tang et al. 2020, Fergus et al. 2023) (Warton et al. 2015, Ovaskainen et al. 2016, Kopp et al.
925 2023). We used a modeling approach to relate major anthropogenic disturbances in the
926 watershed to several in-stream physiochemical variables that influence the distribution of benthic
927 macroinvertebrates (Kopp et al. 2023) and therefore circumvent the need to define background

928 physiochemical conditions using reference sites (Dodds and Oakes 2004, Herlihy and Sifneos
929 2008). Unsurprisingly, anthropogenic disturbance variables were ranked among the most
930 important predictors for nutrients, salinity, and substrate diameter, supporting the notion that
931 these environmental gradients are susceptible to alteration by anthropogenic activities (Paulsen et
932 al. 2008, Lin et al. 2021, Zak et al. 2021, Kaufmann et al. 2022, Kaushal et al. 2023, Sabo et al.
933 2023). Overall, our results suggest that nutrient and salinity concentrations could be lower and
934 substrate diameter could be coarser if anthropogenic disturbances were removed.

935 The proportion of sites where we inferred altered physiochemical conditions varied
936 among the ecoregions and could indicate regional differences in the extent and magnitude of
937 specific anthropogenic activities, as well as the types of disturbance and their effects on their
938 physiochemical and biotic conditions. For example, the larger portion of sites with higher than
939 expected total nitrogen in Temperate Plains and Upper Midwest suggests that nutrient inputs
940 associated with agricultural activities may be elevating in-stream nutrient concentrations.
941 Although, this feature kept our model sufficiently tractable it limits our ability to make accurate
942 predictions to new locations (Abrego and Ovaskainen 2023). However, predicting to new
943 locations was not our objective. Rather, our models establish empirical relationships between
944 macroinvertebrate occurrences and then evaluate how those assemblages might differ if select
945 physiochemical conditions changed while all else remained constant. Furthermore, because the
946 data we used for our analysis was collected as part of a probabilistic survey, designed to be
947 representative of the population of stream and rivers (Olsen and Peck 2008), our site-specific
948 inferences can be aggregated to elucidate regional and sub-continental patterns. Indeed, it may be
949 undesirable to refit models every time an assessment is needed, and future efforts should focus

950 on improving predictive abilities by including immutable factors in addition to stressor gradients
951 (Yuan et al. 2024).

952 Importantly our approach differs significantly from traditional, reference site-based
953 approaches (Lin et al. 2021, Sabo et al. 2023)(Hawkins et al. 2000, Herlihy et al. 2008, USEPA
954 2023). Similarly, chloride concentrations exceed hindcاستimated estimates for the majority of streams
955 in Northern Appalachians, potentially reflecting the contribution of road salts in freshwater
956 salinization (Kaushal et al. 2023). The Temperate Plains had a relatively large proportion of sites
957 with finer substrates and reduced natural riparian vegetation and emphasizes the importance of
958 maintaining native vegetation in riparian areas (Matthews 1988, Dodds et al. 2004). Finally,
959 regional sulfate concentrations in streams can be influenced by acid mine drainage, fertilizer
960 leaching, and industrial wastewater inputs (Herlihy et al. 1990, Zak et al. 2021). We found that
961 approximately 10% of sites in the Southern Appalachians and Upper Midwest could have lower
962 concentrations of sulfate, possibly implicating these activities in elevating sulfate concentrations.
963 The difference between present-day and 1940–1950 climate averages were generally less than 2
964 standard deviations, indicating similarities in mean summer temperature and total annual
965 precipitation. However, averages may not capture changes in variability or extreme events and
966 could potentially oversimplify several other important aspects of climatic change that have
967 occurred in the United States (USEPA 2024). Importantly, using landscape variables to predict
968 in-stream physiochemical provides a more direct linkage between human disturbance and
969 biological assemblages than using landcover variables alone (Kilgour and Stanfield 2005, Yuan
970 et al. 2024).

971 . Foremost, our efforts focus explicitly on a relatively small number of potential stressors
972 whereas reference-site approaches focus implicitly on a theoretically larger, but undefined

973 number of stressors. For example, the difference between a test site and reference sites could be
974 related to a number of other stressors that were not used as biotic screens but co-occur with them
975 (Herlihy et al. 2008, Paulsen et al. 2008). Alternatively, in our model-based approach,
976 differences between present day and hindcast assemblages are only related to changes in the
977 physiochemical variables included as fixed effects. Although the former may elucidate general
978 disturbance effects, it is difficult to attribute differences between assemblages to a specific
979 environmental disturbance (Paulsen et al. 2008). On the other hand, our model-based approach
980 enhances interpretations with respect to specific physiochemical variables but may omit
981 important anthropogenic stressors that were not explicitly included in the model. Thus, the
982 choice between reference site-based and model-based assessments may be contingent on whether
983 the environmental gradients that are commonly disturbed by human activities can be identified
984 and appropriately modeled.

985 The JSMDs were also fitted using bayesian inference and, as such, yielded posterior
986 distributions that can be used for hypothesis testing (Johnson et al. 2022). Specifically, we
987 evaluate whether the effects of human disturbance on physiochemical variables were sufficient to
988 alter benthic assemblages on a site-specific basis. In contrast, traditional reference site-based
989 approaches evaluate each test site based on quantiles of the distribution obtained from reference
990 sites (Herlihy et al. 2008, USEPA 2023). In this regard, JSMDs and bayesian inference may be
991 advantageous in interpreting biological condition estimates because they account for
992 uncertainty in predictions for each site. Furthermore, we also included all genera from the
993 regional taxa pool while some typical reference site-based approaches include only taxa that
994 occur at reference sites. This implies that in the absence of human disturbance those taxa should
995 occur at all locations regardless of other taxa. This is problematic because genera that tolerate

996 anthropogenic disturbance probably evolved under similar conditions that occurred naturally or
997 have remarkable plasticity (Wiens et al. 2010, Heino et al. 2013). Finally, our model-based
998 approach can reveal compositional changes. Although the probabilistic adaptation of Jaccard's
999 similarity calculation suggests that predicting individual genera is more challenging than
1000 aggregated genus richness, this analysis provided additional information that is generally not
1001 easily available from other approaches (Hawkins 2006, Van Sickle 2008) and enhanced our
1002 ability to identify sites that are potentially affected by human disturbance.

1003 Selecting the appropriate taxonomic level that specimens should be identified to is a
1004 critical decision in biological assessment (Chessman et al. 2007). We focused exclusively on
1005 taxa that were identified to the genus level and excluded those that could not be unambiguously
1006 identified. This possibly increases false absences rates in our study, but NRSA identifies most
1007 organisms to their genus such that these instances are relatively rare. Further, common taxonomy
1008 alleviates some ambiguity associated with aggregating unresolved taxa into operational
1009 taxonomic units (Yuan et al. 2008) and improves transferability of our genus-environment
1010 relationships to other studies. Fixed-count subsampling, performed as part of the standardized
1011 NRSA protocol, could also increase false absences in our study. Although, explicitly modeling
1012 taxon-specific detection probabilities from replicate subsamples may be an interesting avenue for
1013 future research (Doser et al. 2023, Doser et al. 2024), these data are presently unavailable.
1014 Finally, focusing on genera could also reduce interspecific variation, but species-level taxonomy
1015 was not available for this dataset. In general, finer taxonomic resolution would be substantially
1016 more expensive and perhaps only yield marginal benefits (Chessman et al. 2007). Nonetheless,
1017 metabarcoding approaches may have shown promise for bioassessments (Smucker et al. 2024)

1018 and could avoid some of the limitations associated with selecting the appropriate taxonomic
1019 resolution.

1020 Although the model-based approaches may have some advantages over reference site-
1021 based approaches, they require further investigation before they can be fully assimilated into
1022 biomonitoring programs. Nevertheless, the National Rivers and Streams Assessment reports
1023 biological condition estimates at for the 2018-19 survey using a multi-metric index (USEPA
1024 2024b) and it is worthwhile to compare our results to those. The results from NRSA are available
1025 at <https://riverstreamassessment.epa.gov/dashboard>. Based on a ranked comparison, the largest
1026 disagreements pertained to the percentage of stream miles in poor condition in the Plains
1027 ecoregions. Specifically, our model-based approach showed more streams to be altered in the
1028 Northern Plains, Southern Plains and Temperate Plains. Indeed, the US Great Plains (i.e.
1029 Temperate Plains, Northern Plains and Southern Plains) have undergone extensive conversion
1030 from grasslands to agriculture such that there may be few sites that are undisturbed (Samson and
1031 Knopf 1994, Dodds et al. 2004, Olimb and Robinson 2019). Because of the large extent of
1032 anthropogenic activities in these regions, it is likely that the reference sites used to assess
1033 biological condition are potentially lower quality (Herlihy et al. 2008). Conversely, NRSA
1034 ranked the Coastal Plains as having a larger percentage of streams in poor condition than our
1035 model-based approach. Unfortunately, because of potential extrapolation associated with our
1036 model-based assessment, the Coastal Plains also had a relatively large number of streams that
1037 could not be assessed. The rankings for the remaining ecoregions tended be similar and both
1038 methods agreed that streams in the Western Mountains were the least-disturbed among the
1039 ecoregions. Indeed, about 75% of the land area of this ecoregion is in federal ownership, which
1040 could convey some protection to streams and rivers (Jenkins et al. 2015). Furthermore, Western

1041 Mountains have a relatively shorter history of anthropogenic disturbance than other ecoregions in
1042 the United States.

1043

1044 *Macroinvertebrate response to anthropogenic disturbance*

1045 Substituting hindcast physiochemical conditions into the JSMDs elucidated how
1046 macroinvertebrate assemblages may change if human disturbance was removed. Hindcast
1047 physiochemical conditions tended to increase estimates of genus richness, suggesting that in
1048 many ecoregions fewer genera persist when anthropogenic disturbances alter physiochemical
1049 conditions. Indeed, other studies that have found either modest increases in macroinvertebrate
1050 richness. Beyond applications for bioassessment, our model-based approach contributes to a
1051 growing literature devoted to understanding of recent trends in macroinvertebrates (Gebert et al.
1052 2022, Rumschlag et al. 2023) (Crossley et al. 2020, Jähnig et al. 2021, Gebert et al. 2022, Spake
1053 et al. 2022, Rumschlag et al. 2023) or no clear evidence for a widespread decline of insects
1054 (Crossley et al. 2020). However, few datasets span more than several decades which limits our
1055 ability to assess trends beyond a relatively recent historical time period (Blüthgen et al. 2022)
1056 and could potentially reflect a shifting baseline effect (Humphries and Winemiller 2009, Soga
1057 and Gaston 2018). An advantage of our approach is that space-for-time substitutions leverage
1058 spatial variability to increase the length of the environmental gradient being assessed such that
1059 our pristine scenario represents an earlier period before major anthropogenic disturbances.
1060 Further, the survey design implemented by NRSA reduces confounding spatial and temporal
1061 dimensions that could be present in datasets compiled from multiple sources. In our study,
1062 hindcasting estimates tended to show increases in genus richness, suggesting that in many
1063 ecoregions fewer genera persist when physiochemical conditions are disturbed. Other studies

1064 have reported either modest increases in macroinvertebrate richness (Gebert et al. 2022,
1065 Rumschlag et al. 2023) or no clear evidence for a widespread decline of insects (Crossley et al.
1066 2020). Because few datasets span more than several decades, an advantage of our space-for-time
1067 approach is that it could reflect a period before major anthropogenic disturbances (Blüthgen et al.
1068 2022). In addition, the survey design and consistent methodology implemented by NRSA
1069 reduces the potential for confounding spatial and temporal dimensions that could be present in
1070 datasets that were compiled from multiple sources (Jähnig et al. 2021, (Jähnig et al. 2021,
1071 Blüthgen et al. 2022, Boyd et al. 2023), Boyd et al. 2023). Nonetheless, although our results
1072 suggest that sites typically support fewer macroinvertebrate genera under present-day conditions,
1073 we did detect increased in genus richness at several sites in the Northern Plains and Western
1074 Mountains which suggests that macroinvertebrate response to disturbance could be context
1075 dependent (Powell et al. 2023)(Powell et al. 2023).

1076 We compared genus-specific occurrence probabilities under present-day and hindcasted
1077 conditions to identify which genera contribute to assemblage level patterns. Insects in most
1078 ecoregions tended to decrease with disturbance while non-insects tended to increase. To
1079 elucidate whichThis suggests that insect genera may, in general, be less tolerant to anthropogenic
1080 disturbance (Jähnig et al. 2021) and aligns with other studies that have documented declines in
1081 insect richness (Rumschlag et al. 2023). Among insects, Ephemeroptera, Plecoptera and
1082 Trichoptera (EPT) are often used as indicators of anthropogenic disturbance (Stoddard et al.
1083 2008) and may have already lost a considerable proportion of species (Sánchez-Bayo and
1084 Wyckhuys 2019). This is especially concerning because we found that EPT occurrence
1085 probabilities could continue to decrease in many regions of the US. Among non-insects, non-
1086 arthropods included members of Annelida, Nemertea, and Platyhelminthes tended to increase at

1087 a relatively large proportion of sites which is consistent with high pollution tolerance values
1088 typically assigned to these taxa (Carlisle et al. 2007, Griffith 2023).

1089 Because our approach was based on quantifying genus-environment relationships along
1090 environmental gradients that are typically altered by human activity, we could also compare the
1091 relative effects of changing specific categories of physiochemical variables were most important
1092 in structuring on macroinvertebrate assemblages, we created scenarios by changing either NTL
1093 and PTL (Nutrient Scenario), CL and SO4 (Salinity Scenario), RPDI and SUBD (Habitat
1094 Scenario), or MSAT and TPRCP (Climate Scenario). The increase in estimated. For example,
1095 higher genus richness under present-day conditions in the Northern Plains and Western
1096 Mountains ecoregions was also present in the compared to hindcasted habitat scenario
1097 and variables suggests that disturbance may have increased taxa richness by disturbing the
1098 riparian area or substrate diameter. Riparian disturbance was present at majority of these sites
1099 and the index we used to measure it consists of several direct measures of human disturbance
1100 (USEPA 2017a). Cattle genera richness. Interestingly, cattle in pasture and rangelands were one
1101 of among the most common types of disturbance measured by riparian disturbance index RPDI for
1102 these regions this region (USEPA 2023)(USEPA 2023) and could potentially increase taxagenera
1103 richness by suppressing riparian forest cover and, in turn, elevating primary production
1104 (Mittelbach et al. 2001, Tonkin et al. 2013)(Mittelbach et al. 2001, Tonkin et al. 2013).
1105 Conversely, in the Northern Appalachians and Coastal Plains, where the presence of trash,
1106 landfills or buildings were the most common factors contributing to the riparian disturbance
1107 index for RPDI, we found that genus richness was presently lower which is consistent with the
1108 negative effects of urbanization on stream macroinvertebrate assemblages (Morse et al.
1109 2003).(Morse et al. 2003). We also found different effects with respect to changes in

1110 precipitation whereby genus richness increased in the Northern Plains and decreased in Southern
1111 Appalachians. Collectively, these results elucidate patterns reveal which environmental
1112 variables couldmay be most important forin structuring microinvertebrate assemblages and in
1113 what contexts-

1114 ~~We compared taxon specific occurrence probabilities under present-day and hindcasted~~
1115 ~~conditions to identify taxa that were considered to increase or decrease under anthropogenic~~
1116 ~~disturbance. Insects in most ecoregions tended to decrease with disturbance while non-insects~~
1117 ~~tended to increase. This suggests that insect taxa may, in general, be comparatively less tolerant~~
1118 ~~to anthropogenic disturbance (Jähnig et al. 2021) and aligns with other studies that have~~
1119 ~~documented declines in insect richness (Rumschlag et al. 2023). Among insects, Ephemeroptera,~~
1120 ~~Plecoptera and Trichoptera (EPT) are often used as indicators of anthropogenic disturbance~~
1121 ~~(Stoddard et al. 2008) and considered to have already lost a considerable proportion of species~~
1122 ~~(Sánchez-Bayo and Wyckhuys 2019).~~ We found that the occurrence probabilities for the EPT
1123 ~~taxa – an interpretation~~ that ~~still exist could decrease at approximately 25% of sites in the~~
1124 ~~Temperate Plains. Indeed, these taxa may~~would not be especially sensitive to nutrient additions
1125 ~~(Nessel et al. 2023)~~ and the nutrient scenario we created here resulted in decreased richness at
1126 ~~majority of these sites without removing disturbance from any other environmental gradient.~~
1127 ~~Among non-insects, non-arthropods included members of Annelida, Nemertea, and~~
1128 ~~Platyhelminthes tended to increase at a relatively large proportion of sites which is consistent~~
1129 ~~with high pollution tolerance values typically assigned to these taxa (Carlisle et al. 2007, Griffith~~
1130 ~~2023).~~

1131 ~~The proportion of sites that changed in richness or composition could reveal differences~~
1132 ~~among ecoregions in the effects of anthropogenic disturbance on macroinvertebrate assemblages.~~

1133 We found disturbance effected macroinvertebrate assemblages at the greatest proportion of sites
1134 in the Northern Plains, Southern Plains and Temperate Plains compared to other regions. The US
1135 Great Plains have undergone extensive conversion from grasslands to agriculture such that there
1136 may be few sites that are undisturbed (Samson and Knopf 1994, Dodds et al. 2004, Olimb and
1137 Robinson 2019). Conversely, we found that present-day assemblages in the Western Mountains
1138 were generally similar to hindcast predictions. About 75% of the land area of this ecoregion is
1139 classified as federal ownership land which could convey some protection to streams and rivers
1140 (Jenkins et al. 2015). Further, this region has a relatively short history of anthropogenic
1141 disturbance compared to other regions of the United States, such that there may be a relatively
1142 large number of streams and rivers that are undisturbed. Our JSDM approach also allowed us to
1143 assess compositional change in addition to changes in richness resulting from disturbance. We
1144 found that assemblage composition at some sites in all regions could be affected by
1145 anthropogenic disturbance without a subsequent change in richness (Van Sickle 2008), which
1146 suggests that tolerant taxa may have replaced sensitive taxa at many sites (Jähnig et al. 2021,
1147 Rumschlag et al. 2023).

1148

1149 *Application for biological assessment without reference sites*

1150 Identifying minimally disturbed reference sites to assess the effects of anthropogenic
1151 disturbance on biological assemblages is challenging because there are few existing sites in this
1152 condition (Stoddard et al. 2006). Least disturbed is often adopted as a suitable alternative but is
1153 difficult to consistently define, vary in quality, and can be spatially aggregated (Hawkins et al.
1154 2010, McNellie et al. 2020). These challenges have motivated several efforts to explore
1155 alternative approaches for assessing biological condition without relying on reference sites

1156 (Chessman and Royal 2004, Kilgour and Stanfield 2005, Elias et al. 2016, Yuan et al. 2024).
1157 Beyond elucidating the potential effects of anthropogenic disturbance on macroinvertebrate
1158 assemblages, our approach complements these efforts by developing a two-stage modeling
1159 approach that combines random forest predictions of physiochemical conditions in the absence
1160 of anthropogenic disturbance possible with taxon-environment relationships quantified using joint
1161 species distribution modeling.

1162 Specifically, our efforts include the entire regional taxa pool and incorporated taxon-
1163 specific responses to physiochemical variables that are often altered by human activities. This is
1164 consistent with Yuan et al. (2024) but differs from traditional approaches that quantify
1165 relationships between taxa that occur at least disturbed reference sites and natural (immutable)
1166 environmental gradients (e.g. watershed area, elevation, lithology) (Hawkins et al. 2000, Clarke
1167 et al. 2003, Hallstan et al. 2012). By focusing on physiochemical variables that are commonly
1168 altered by human disturbance, we can more directly infer the consequences of disturbance, or the
1169 potential effectiveness of efforts to restore or remediate specific stressors, on macroinvertebrate
1170 assemblages. We also used site-level random effects to statistically control for unmeasured
1171 environmental variables or biotic interactions using JSMDs (Kopp et al. 2023). This had the
1172 added benefit of keeping our model sufficiently tractable but assumes that only the
1173 environmental variables we included in our analysis act as agents of change for the assemblage
1174 and limits our ability to make accurate predictions to new locations. Indeed the data we used for
1175 our analysis were collected as part of a probabilistic survey specifically designed to be
1176 representative of the population of stream and rivers (Olsen and Peck 2008) thus our site-specific
1177 analysis should capture general patterns across CONUS. Nonetheless future efforts to develop

1178 ~~these models for biological assessment purposes will benefit from improving the predictive~~
1179 ~~capability of these models~~reference site-based approaches.

1180

1181

1182 *Conclusions*

1183 We used a ~~two-stage modeling process~~model-based approach to assess the potential
1184 effects of anthropogenic disturbance on physiochemical gradients and benthic macroinvertebrate
1185 assemblages. Our approach combines ~~taxon~~genus-environment relationships with estimates of ~~a~~
1186 number of important dimensions in physiochemical ~~conditions~~condition after removing
1187 anthropogenic disturbance. ~~Increasingly~~The number of sites in minimally disturbed condition are
1188 ~~disappearing and progressively diminishing~~, so methods that circumvent the need for reference
1189 sites ~~in the~~for biological ~~assessment~~assessments of streams and rivers are crucial to
1190 understanding the extent of anthropogenic impacts. Importantly, ~~this~~our framework could
1191 provide an avenue to conduct biological assessment without depending on least disturbed
1192 reference sites.

1193

1194

1195 Acknowledgements

1196 We thank S. Paulsen, L. Riato, E. Fergus, K. Blocksom, and D. Peck for valuable
1197 discussions throughout the development of this project. D. Benkendorf ~~and~~, S. Rumschlag, D.
1198 Carlisle and an anonymous reviewer provided comments on earlier versions of this manuscript.
1199 The information in this document has been funded entirely by the U.S. Environmental Protection
1200 Agency, in part through appointments to the Internship/Research Participation Program at the

1201 Office of Research and Development, U.S. Environmental Protection Agency, administered by
1202 the Oak Ridge Institute for Science and Education through an interagency agreement. The views
1203 expressed in this article are those of the authors and do not necessarily represent the views or
1204 policies of the U.S. Environmental Protection Agency. Any mention of trade names, products or
1205 services does not imply an endorsement by the U.S. Government or the U.S. Environmental
1206 Protection Agency. The EPA does not endorse commercial products, services, or enterprises.

For Review Only

1207 enterprise

1208

1209 References:

1210 Bailey, R. C., S. Linke, and A. G. Yates. 2014. Bioassessment of freshwater ecosystems using the
1211 Reference Condition Approach: comparing established and new methods with common data
1212 sets. *Freshwater Science* **33**:1204-1211.

1213 Abrego, N., and O. Ovaskainen. 2023. Evaluating the predictive performance of presence-
1214 absence models: Why can the same model appear excellent or poor? *Ecology and*
1215 *Evolution* **13**:e10784.

1216 Blüthgen, N., M. Staab, R. Achury, and W. W. Weisser. 2022. Unravelling insect declines: can
1217 space replace time? *Biology Letters* **18**:20210666.

1218 Bouleau, G., and D. Pont. 2015. Did you say reference conditions? Ecological and socio-
1219 economic perspectives on the European Water Framework Directive. *Environmental*
1220 *Science & Policy* **47**:32-41.

1221 Boyd, R. J., G. D. Powney, and O. L. Pescott. 2023. We need to talk about nonprobability
1222 samples. *Trends in Ecology & Evolution* **38**:521-531.

1223 Breiman, L. 2001. Random Forests. *Machine Learning* **45**:5-32.

1224 Buss, D. F., D. M. Carlisle, T.-S. Chon, J. Culp, J. S. Harding, H. E. Keizer-Vlek, W. A.
1225 Robinson, S. Strachan, C. Thirion, and R. M. Hughes. 2014. Stream biomonitoring using
1226 macroinvertebrates around the globe: a comparison of large-scale programs.
1227 *Environmental Monitoring and Assessment* **187**:4132.

1228 Calabrese, J. M., G. Certain, C. Kraan, and C. F. Dormann. 2014. Stacking species distribution
1229 models and adjusting bias by linking them to macroecological models. *Global Ecology*
1230 and Biogeography

and Biogeography **23**:99-112.

- 1231 Carlisle, D. M., M. R. Meador, S. R. Moulton, and P. M. Ruhl. 2007. Estimation and application
1232 of indicator values for common macroinvertebrate genera and families of the United
1233 States. *Ecological Indicators* 7:22-33.
- 1234 ~~Chen, K., J. R. Olson, J. J. Vander Laan, R. A. Hill, B. Wang, and C. P. Hawkins. 2019. Improving the
1235 performance of ecological indices by balancing reference site quality and representativeness.
1236 Hydrobiologia~~ **837**:177-194.
- 1237 Chessman, B., S. Williams, and C. Besley. 2007. Bioassessment of streams with
1238 macroinvertebrates: effect of sampled habitat and taxonomic resolution. Journal of the
1239 North American Benthological Society **26**:546-565.
- 1240 Chessman, B. C., and M. J. Royal. 2004. Bioassessment without reference sites: use of
1241 environmental filters to predict natural assemblages of river macroinvertebrates. *Journal
1242 of the North American Benthological Society* **23**:599-615.
- 1243 ~~Clarke, R. T., J. F. Wright, and M. T. Furse. 2003. RIVPACS models for predicting the expected
1244 macroinvertebrate fauna and assessing the ecological quality of rivers. Ecological Modelling
1245~~ **160**:219-233.
- 1246 Crossley, M. S., A. R. Meier, E. M. Baldwin, L. L. Berry, L. C. Crenshaw, G. L. Hartman, D.
1247 Lagos-Kutz, D. H. Nichols, K. Patel, S. Varriano, W. E. Snyder, and M. D. Moran. 2020.
1248 No net insect abundance and diversity declines across US Long Term Ecological
1249 Research sites. *Nature Ecology & Evolution* **4**:1368-1376.
- 1250 Deflem, I. S., E. Bennetsen, Ø. H. Opdal, F. C. F. Calboli, O. Ovaskainen, G. Van Thuyne, F.
1251 A. M. Volckaert, and J. A. M. Raeymaekers. 2021. Predicting fish community responses
1252 to environmental policy targets. *Biodiversity and Conservation* **30**:1457-1478.
- 1253 Dodds, W. K., K. Gido, M. R. Whiles, K. M. Fritz, and W. J. Matthews. 2004. Life on the Edge:
1254 The Ecology of Great Plains Prairie Streams. *BioScience* **54**:205-216.
- 1255 Dodds, W. K., and R. M. Oakes. 2004. A technique for establishing reference nutrient
1256 concentrations across watersheds affected by humans. *Limnology and Oceanography:
1257 Methods* **2**:333-341.

- 1258 Doser, J. W., A. O. Finley, and S. Banerjee. 2023. Joint species distribution models with
1259 imperfect detection for high-dimensional spatial data. *Ecology* **104**:e4137.
- 1260 Doser, J. W., A. O. Finley, M. Kéry, and E. F. Zipkin. 2024. spAbundance: An R package for
1261 single-species and multi-species spatially explicit abundance models. *Methods in
1262 Ecology and Evolution* n/a.
- 1263 Dudgeon, D., A. H. Arthington, M. O. Gessner, Z. Kawabata, D. J. Knowler, C. Lévêque, R. J.
1264 Naiman, A. H. Prieur-Richard, D. Soto, M. L. Stiassny, and C. A. Sullivan. 2006.
1265 Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol
1266 Rev Camb Philos Soc* **81**:163-182.
- 1267 Dumelle, M., T. Kincaid, A. R. Olsen, and M. Weber. 2023. spsurvey: Spatial Sampling Design
1268 and Analysis in R. *Journal of Statistical Software* **105**:1 - 29.
- 1269 Elias, C. L., A. R. Calapez, S. F. P. Almeida, B. Chessman, N. Simões, and M. J. Feio. 2016.
1270 Predicting reference conditions for river bioassessment by incorporating boosted trees in
1271 the environmental filters method. *Ecological Indicators* **69**:239-251.
- 1272 Elith, J., and J. R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and
1273 Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and
1274 Systematics* **40**:677-697.
- 1275 Fergus, C. E., J. R. Brooks, P. R. Kaufmann, A. T. Herlihy, R. A. Hill, R. M. Mitchell, and P.
1276 Ringold. 2023. Disentangling natural and anthropogenic effects on benthic
1277 macroinvertebrate assemblages in western US streams. *Ecosphere* **14**:e4688.
- 1278 Franklin, J. 2010. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge
1279 University Press, Cambridge.

- 1280 Gebert, F., M. K. Obrist, R. Siber, F. Altermatt, K. Bollmann, and N. Schuwirth. 2022. Recent
1281 trends in stream macroinvertebrates: warm-adapted and pesticide-tolerant taxa increase in
1282 richness. *Biology Letters* **18**:20210513.
- 1283 Griffith, M. B. 2023. Development of national stressor-specific genus sensitivity values. *Science
1284 of The Total Environment* **895**:165121.
- 1285 ~~Hallstan, S., R. K. Johnson, E. Willén, and U. Grandin. 2012. Comparison of classification then modelling
1286 and species-by-species modelling for predicting lake phytoplankton assemblages. Ecological
1287 Modelling~~ **231**:11-19.
- 1288 Guisan, A., W. Thuiller, and N. E. Zimmermann. 2017. Habitat Suitability and Distribution
1289 Models: With Applications in R. Cambridge University Press, Cambridge.
- 1290 Hawkins, C. P. 2006. QUANTIFYING BIOLOGICAL INTEGRITY BY TAXONOMIC
1291 COMPLETENESS: ITS UTILITY IN REGIONAL AND GLOBAL ASSESSMENTS.
1292 *Ecological Applications* **16**:1277-1294.
- 1293 Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000. DEVELOPMENT AND
1294 EVALUATION OF PREDICTIVE MODELS FOR MEASURING THE BIOLOGICAL
1295 INTEGRITY OF STREAMS. *Ecological Applications* **10**:1456-1477.
- 1296 Hawkins, C. P., J. R. Olson, and R. A. Hill. 2010. The reference condition: predicting
1297 benchmarks for ecological and water-quality assessments. *Journal of the North American
1298 Benthological Society* **29**:312-343.
- 1299 ~~Hering, D., A. Borja, J. Carstensen, L. Carvalho, M. Elliott, C. K. Feld, A. S. Heiskanen, R. K. Johnson, J.
1300 Moe, D. Pont, A. L. Solheim, and W. v. de Bund. 2010. The European Water Framework Directive
1301 at the age of 10: A critical review of the achievements with recommendations for the future.
1302 Science of The Total Environment~~ **408**:4007-4019.
- 1303 Heino, J., D. Schmera, and T. Erős. 2013. A macroecological perspective of trait patterns in
1304 stream communities. *Freshwater Biology* **58**:1539-1555.
- 1305 Herlihy, A. T., P. R. Kaufmann, ~~and~~ M. E. Mitch, ~~and~~ D. D. Brown. 1990. Regional estimates of acid
1306 mine drainage impact on streams in the mid-atlantic and Southeastern. **1991. Stream**

- 1307 chemistry in the eastern United States. *Water, Air, and Soil Pollution* **50**:91-107; 2. Current
1308 sources of acidity in acidic and low acid-neutralizing capacity streams. *Water Resources*
1309 Research **27**:629-642.
- 1310 Herlihy, A. T., S. G. Paulsen, J. V. Sickle, J. L. Stoddard, C. P. Hawkins, and L. L. Yuan. 2008.
1311 Striving for consistency in a national assessment: the challenges of applying a reference-
1312 condition approach at a continental scale. *Journal of the North American Benthological*
1313 *Society* **27**:860-877.
- 1314 Herlihy, A. T., and J. C. Sifneos. 2008. Developing nutrient criteria and classification schemes
1315 for wadeable streams in the conterminous US. *Journal of the North American*
1316 *Benthological Society* **27**:932-948.
- 1317 Herlihy, A. T., J. L. Stoddard, and C. B. Johnson. 1998. The Relationship Between Stream
1318 Chemistry and Watershed Land Cover Data in the Mid-Atlantic Region, U.S. Water, Air,
1319 and Soil Pollution **105**:377-386.
- 1320 Hill, R. A., M. H. Weber, S. G. Leibowitz, A. R. Olsen, and D. J. Thornbrugh. 2016. The
1321 Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the
1322 Conterminous United States. *JAWRA Journal of the American Water Resources*
1323 *Association* **52**:120-128.
- 1324 Hughes, R. M., and D. V. Peck. 2008. Acquiring data for large aquatic resource surveys: the art
1325 of compromise among science, logistics, and reality. *Journal of the North American*
1326 *Benthological Society* **27**:837-859.
- 1327 ~~Humphries, P., and K. O. Winemiller. 2009. Historical Impacts on River Fauna, Shifting Baselines, and
1328 Challenges for Restoration. BioScience~~ **59**:673-684.
- 1329 ~~Jackson, S., and W. Davis. 1994. Meeting the Goal of Biological Integrity in Water Resource Programs in
1330 the US Environmental Protection Agency. Journal of the North American Benthological Society~~
1331 **13**:592-597.

- 1332 Jähnig, S. C., V. Baranov, F. Altermatt, P. Cranston, M. Friedrichs-Manthey, J. Geist, F. He, J.
1333 Heino, D. Hering, F. Hölker, J. Jourdan, G. Kalinkat, J. Kiesel, F. Leese, A. Maasri, M.
1334 T. Monaghan, R. B. Schäfer, K. Tockner, J. D. Tonkin, and S. Domisch. 2021. Revisiting
1335 global trends in freshwater insect biodiversity. *WIREs Water* **8**:e1506.
- 1336 Jenkins, C. N., K. S. Van Houtan, S. L. Pimm, and J. O. Sexton. 2015. US protected lands
1337 mismatch biodiversity priorities. *Proceedings of the National Academy of Sciences*
1338 **112**:5081-5086.
- 1339 Johnson, A. A., M. Q. Ott, and M. Dogucu. 2022. Bayes Rules!: An Introduction to Applied
1340 Bayesian Modeling Chapman and Hall/CRC.
- 1341 Kaufmann, P. R., R. M. Hughes, S. G. Paulsen, D. V. Peck, C. W. Seeliger, T. Kincaid, and R.
1342 M. Mitchell. 2022a. Physical habitat in conterminous US streams and Rivers, part 2: A
1343 quantitative assessment of habitat condition. Ecological Indicators **141**:109047.
- 1344 Kaufmann, P. R., R. M. Hughes, S. G. Paulsen, D. V. Peck, C. W. Seeliger, M. H. Weber, and R.
1345 M. Mitchell. 2022b. Physical habitat in conterminous US streams and rivers, Part 1:
1346 Geoclimatic controls and anthropogenic alteration. Ecological Indicators **141**:109046.
- 1347 Kaufmann, P. R., P. Levine, E.G. Robison, C. Seeliger, and D.V. Peck. 1999. Quantifying
1348 Physical Habitat in Wadeable Streams. EPA/620/R-99/003, U.S. Environmental
1349 Protection Agency, Washington, D.C.
- 1350 Kaushal, S. S., G. E. Likens, P. M. Mayer, R. R. Shatkay, S. A. Shelton, S. B. Grant, R. M. Utz,
1351 A. M. Yaculak, C. M. Maas, J. E. Reimer, S. V. Bhide, J. T. Malin, and M. A. Rippy.
1352 2023. The anthropogenic salt cycle. *Nature Reviews Earth & Environment* **4**:770-784.
- 1353 Kilgour, B., and L. Stanfield. 2005. A new approach for defining reference conditions in streams.
1354 AGU Spring Meeting Abstracts.

- 1355 Kopp, D. A., J. L. Stoddard, R. A. Hill, J. M. Doyle, P. R. Kaufmann, A. T. Herlihy, and S. G.
1356 Paulsen. 2023. Joint species distribution models reveal taxon-specific sensitivities to
1357 potential anthropogenic alteration. *Freshwater Science* **42**:268-284.
- 1358 ~~Lewis, S. L., and M. A. Maslin. 2015. Defining the Anthropocene. *Nature* **519**:171-180.~~
1359 Lin, J., J. E. Compton, R. A. Hill, A. T. Herlihy, R. D. Sabo, J. R. Brooks, M. Weber, B. Pickard,
1360 S. G. Paulsen, and J. L. Stoddard. 2021. Context is Everything: Interacting Inputs and
1361 Landscape Characteristics Control Stream Nitrogen. *Environmental Science &*
1362 *Technology* **55**:7890-7899.
- 1363 Linke, S., R. H. Norris, D. P. Faith, and D. Stockwell. 2005. ANNA: A new prediction method
1364 for bioassessment programs. *Freshwater Biology* **50**:147-158.
- 1365 ~~Matthews, W. J. 1988. North American Prairie Streams as Systems for Ecological Study. *Journal of the*~~
1366 ~~North American Benthological Society~~ **7**:387-409.
- 1367 McNellie, M. J., I. Oliver, J. Dorrough, S. Ferrier, G. Newell, and P. Gibbons. 2020. Reference
1368 state and benchmark concepts for better biodiversity conservation in contemporary
1369 ecosystems. *Global Change Biology* **26**:6702-6714.
- 1370 Meinshausen, N. 2017. quantregForest: Quantile Regression Forests. R package version 1.3-7.
- 1371 Meyer, H., and E. Pebesma. 2021. Predicting into unknown space? Estimating the area of
1372 applicability of spatial prediction models. Methods in Ecology and Evolution **12**:1620-
1373 1633.
- 1374 Mitchell, R. M., A. T. Herlihy, and R. M. Hughes. 2025. Striving for consistency in a national
1375 lake assessment: Defining reference status and littoral macroinvertebrate condition in
1376 lakes across the conterminous United States. Ecological Indicators **170**:112992.
- 1377 Mittelbach, G. G., C. F. Steiner, S. M. Scheiner, K. L. Gross, H. L. Reynolds, R. B. Waide, M.
1378 R. Willig, S. I. Dodson, and L. Gough. 2001. WHAT IS THE OBSERVED

- 1379 RELATIONSHIP BETWEEN SPECIES RICHNESS AND PRODUCTIVITY? Ecology
1380 82:2381-2396.
- 1381 Morse, C. C., A. D. Huryn, and C. Cronan. 2003. Impervious Surface Area as a Predictor of the
1382 Effects of Urbanization on Stream Insect Communities in Maine, U.S.A. Environmental
1383 Monitoring and Assessment 89:95-127.
- 1384 ~~Nessel, M. P., T. Konnovitch, G. Q. Romero, and A. L. González. 2023. Decline of insects and arachnids~~
1385 ~~driven by nutrient enrichment: A meta-analysis. Ecology 104:e3897.~~
- 1386 O'brien, R. M. 2007. A Caution Regarding Rules of Thumb for Variance Inflation Factors.
1387 Quality & Quantity 41:673-690.
- 1388 Olimb, S. K., and B. Robinson. 2019. Grass to grain: Probabilistic modeling of agricultural
1389 conversion in the North American Great Plains. Ecological Indicators 102:237-245.
- 1390 Olsen, A. R., and D. V. Peck. 2008. Survey design and extent estimates for the Wadeable
1391 Streams Assessment. Journal of the North American Benthological Society 27:822-836.
- 1392 Olson, J. R., and S. M. Cormier. 2019. Modeling Spatial and Temporal Variation in Natural
1393 Background Specific Conductivity. Environmental Science & Technology 53:4316-4325.
- 1394 Olson, J. R., and C. Hawkins, P. . 2013. Developing site-specific nutrient criteria from empirical
1395 models. Freshwater Science 32:719-740.
- 1396 Olson, J. R., and C. P. Hawkins. 2012. Predicting natural base-flow stream water chemistry in
1397 the western United States. Water Resources Research 48.
- 1398 Outhwaite, C. L., P. McCann, and T. Newbold. 2022. Agriculture and climate change are
1399 reshaping insect biodiversity worldwide. Nature 605:97-102.
- 1400 Ovaskainen, O., and N. Abrego. 2020. Joint Species Distribution Modelling: With Applications
1401 in R. Cambridge University Press, Cambridge.

- 1402 [Ovaskainen, O., D. B. Roy, R. Fox, and B. J. Anderson. 2016. Uncovering hidden spatial](#)
1403 [structure in species communities with spatially explicit joint species distribution models.](#)
- 1404 [Methods in Ecology and Evolution 7:428-436.](#)
- 1405 [Ovaskainen, O., G. Tikhonov, A. Norberg, F. Guillaume Blanchet, L. Duan, D. Dunson, T.](#)
1406 [Roslin, and N. Abrego. 2017. How to make more out of community data? A conceptual](#)
1407 [framework and its implementation as models and software. Ecology Letters 20:561-576.](#)
- 1408 Paulsen, S. G., A. Mayio, D. V. Peck, J. L. Stoddard, E. Tarquinio, S. M. Holdsworth, J. V.
1409 Sickle, L. L. Yuan, C. P. Hawkins, A. T. Herlihy, P. R. Kaufmann, M. T. Barbour, D. P.
1410 Larsen, and A. R. Olsen. 2008. Condition of stream ecosystems in the US: an overview of
1411 the first national assessment. Journal of the North American Benthological Society
1412 27:812-821.
- 1413 [Peters, J., B. D. Baets, N. E. C. Verhoest, R. Samson, S. Degroeve, P. D. Becker, and W.](#)
1414 [Huybrechts. 2007. Random forests as a tool for ecohydrological distribution modelling.](#)
- 1415 [Ecological Modelling 207:304-318.](#)
- 1416 Powell, K. E., T. H. Oliver, T. Johns, M. González-Suárez, J. England, and D. B. Roy. 2023.
1417 Abundance trends for river macroinvertebrates vary across taxa, trophic group and river
1418 typology. Glob Chang Biol 29:1282-1295.
- 1419 [Prasad, A. M., L. R. Iverson, and A. Liaw. 2006. Newer Classification and Regression Tree](#)
1420 [Techniques: Bagging and Random Forests for Ecological Prediction. Ecosystems 9:181-](#)
1421 [199.](#)
- 1422 Reynoldson, T. B., R. H. Norris, V. H. Resh, K. E. Day, and D. M. Rosenberg. 1997. The
1423 Reference Condition: A Comparison of Multimetric and Multivariate Approaches to

- 1424 Assess Water-Quality Impairment Using Benthic Macroinvertebrates. Journal of the
1425 North American Benthological Society **16**:833-852.
- 1426 Rumschlag, S. L., M. B. Mahon, D. K. Jones, W. Battaglin, J. Behrens, E. S. Bernhardt, P.
1427 Bradley, E. Brown, F. De Laender, R. Hill, S. Kunz, S. Lee, E. Rosi, R. Schäfer, T. S.
1428 Schmidt, M. Simonin, K. Smalling, K. Voss, and J. R. Rohr. 2023. Density declines,
1429 richness increases, and composition shifts in stream macroinvertebrates. Science
1430 Advances **9**:eadf4896.
- 1431 Sabo, R. D., [C. M. Clark, J. Bash, D. Sobota, E. Cooter, J. P. Dobrowolski, B. Z. Houlton, A.](#)
1432 [Rea, D. Schwede, S. L. Morford, and J. E. Compton. 2019. Decadal Shift in Nitrogen](#)
1433 [Inputs and Fluxes Across the Contiguous United States: 2002–2012. Journal of](#)
1434 [Geophysical Research: Biogeosciences 124](#):3104-3124.
- 1435 [Sabo, R. D., C. M. Clark, D. A. Gibbs, G. S. Metson, M. J. Todd, S. D. LeDuc, D. Greiner, M.](#)
1436 [M. Fry, R. Polinsky, Q. Yang, H. Tian, and J. E. Compton. 2021. Phosphorus Inventory](#)
1437 [for the Conterminous United States \(2002–2012\). Journal of Geophysical Research:](#)
1438 [Biogeosciences 126](#):e2020JG005684.
- 1439 [Sabo, R. D., B. Pickard, J. Lin, B. Washington, C. M. Clark, J. E. Compton, M. Pennino, B.](#)
1440 [Bierwagen, S. D. LeDuc, J. N. Carleton, M. Weber, M. Fry, R. Hill, S. Paulsen, A.](#)
1441 [Herlihy, and J. L. Stoddard. 2023. Comparing Drivers of Spatial Variability in U.S. Lake](#)
1442 [and Stream Phosphorus Concentrations. Journal of Geophysical Research:](#)
1443 [Biogeosciences 128](#):e2022JG007227.
- 1444 Samson, F., and F. Knopf. 1994. Prairie conservation in North America. BioScience **44**:418-421.
- 1445 Sánchez-Bayo, F., and K. A. G. Wyckhuys. 2019. Worldwide decline of the entomofauna: A
1446 review of its drivers. Biological Conservation **232**:8-27.

- 1447 Scherrer, D., H. K. Mod, and A. Guisan. 2020. How to evaluate community predictions without
1448 thresholding? *Methods in Ecology and Evolution* **11**:51-63.
- 1449 Soga, M., and K. J. Gaston. 2018. Shifting baseline syndrome: causes, consequences, and implications.
1450 Frontiers in Ecology and the Environment **16**:222-230.
- 1451 Smucker, N. J., E. M. Pilgrim, C. T. Nielch, L. Gains-Germain, C. Carpenter, J. A. Darling, L. L.
1452 Yuan, R. M. Mitchell, and A. I. Pollard. 2024. Using DNA metabarcoding to characterize
1453 national scale diatom-environment relationships and to develop indicators in streams and
1454 rivers of the United States. Science of The Total Environment **939**:173502.
- 1455 Soranno, P. A., T. Wagner, S. L. Martin, C. McLean, L. N. Novitski, C. D. Provence, and A. R.
1456 Rober. 2011. Quantifying regional reference conditions for freshwater ecosystem
1457 management: A comparison of approaches and future research needs. *Lake and Reservoir*
1458 *Management* **27**:138-148.
- 1459 Spake, R., M. P. Barajas-Barbosa, S. A. Blowes, D. E. Bowler, C. T. Callaghan, M. Garbowksi,
1460 S. D. Jurburg, R. van Klink, L. Korell, E. Ladouceur, R. Rozzi, D. S. Viana, W.-B. Xu,
1461 and J. M. Chase. 2022. Detecting Thresholds of Ecological Change in the Anthropocene.
1462 Annual Review of Environment and Resources **47**:797-821.
- 1463 Stoddard, J. L., A. T. Herlihy, D. V. Peck, R. M. Hughes, T. R. Whittier, and E. Tarquinio. 2008.
1464 A process for creating multimetric indices for large-scale aquatic surveys. *Journal of the*
1465 *North American Benthological Society* **27**:878-891.
- 1466 Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. SETTING
1467 EXPECTATIONS FOR THE ECOLOGICAL CONDITION OF STREAMS: THE
1468 CONCEPT OF REFERENCE CONDITION. *Ecological Applications* **16**:1267-1276.
- 1469 Tang, T., R. J. Stevenson, and J. B. Grace. 2020. The importance of natural versus human factors
1470 for ecological conditions of streams and rivers. *Science of The Total Environment*
1471 **704**:135268.

- 1472 Tikhonov, G., Ø. H. Opedal, N. Abrego, A. Lehikoinen, M. M. J. de Jonge, J. Oksanen, and O.
1473 Ovaskainen. 2020. Joint species distribution modelling with the r-package Hmsc.
1474 Methods in Ecology and Evolution **11**:442-447.
- 1475 Tonkin, J. D., R. G. Death, and K. J. Collier. 2013. Do productivity and disturbance interact to
1476 modulate macroinvertebrate diversity in streams? Hydrobiologia **701**:159-172.
- 1477 USEPA. 2017a. National Rivers and Streams Assessment 2018-19: Field Operations Manual–
1478 Wadeable. EPA-841-B-17-003a. U.S. Environmental Protection Agency, Office of
1479 Water, Washington, DC. .
- 1480 USEPA. 2017b. National Rivers and Streams Assessment 2018-19: Laboratory Operations
1481 Manual. EPA841-B-17-004. U.S. Environmental Protection Agency, Office of Water,
1482 Washington, DC.
- 1483 USEPA. 2023. National Rivers and Streams Assessment 2018-2019 Technical Support
1484 Document. EPA 841-R-22-005. U.S. Environmental Protection Agency, Office of Water
1485 and Office of Research and Development, Washington, D.C. .
- 1486 USEPA. 2024a. Climate change indicators in the United States. EPA 430-R-24-003.
- 1487 USEPA. 2024b. National Rivers and Streams Assessment: The Third Collaborative Survey. U.S.
1488 Environmental Protection Agency, Office of Water and Office of Research and
1489 Development.
- 1490 Van Sickle, J. 2008. An index of compositional dissimilarity between observed and expected
1491 assemblages. Journal of the North American Benthological Society **27**:227-235.
- 1492 Vörösmarty, C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S.
1493 Glidden, S. E. Bunn, C. A. Sullivan, C. R. Liermann, and P. M. Davies. 2010. Global
1494 threats to human water security and river biodiversity. Nature **467**:555-561.

- 1495 Warton, D. I., F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F.
1496 K. C. Hui. 2015. So Many Variables: Joint Modeling in Community Ecology. *Trends in
1497 Ecology & Evolution* **30**:766-779.
- 1498 Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I.
1499 Damschen, T. Jonathan Davies, J.-A. Grytnes, S. P. Harrison, B. A. Hawkins, R. D. Holt,
1500 C. M. McCain, and P. R. Stephens. 2010. Niche conservatism as an emerging principle in
1501 ecology and conservation biology. *Ecology Letters* **13**:1310-1324.
- 1502 Wilkinson, D. P., N. Golding, G. Guillera-Arroita, R. Tingley, and M. A. McCarthy. 2021.
1503 Defining and evaluating predictions of joint species distribution models. *Methods in
1504 Ecology and Evolution* **12**:394-404.
- 1505 ~~Yuan, L. L. 2006. Theoretical predictions of observed to expected ratios in RIVPACS-type predictive
1506 model assessments of stream biological condition. *Journal of the North American Benthological
1507 Society* **25**:841-850.~~
- 1508 Wood, G. 2008. Thresholds and criteria for evaluating and communicating impact significance in
1509 environmental statements: ‘See no evil, hear no evil, speak no evil’? *Environmental
1510 Impact Assessment Review* **28**:22-38.
- 1511 Yuan, L. L., D. M. Carlisle, R. M. Mitchell, and A. I. Pollard. 2024. Using models of local
1512 environmental conditions for biological assessment. *Freshwater Science*:000-000.
- 1513 Yuan, L. L., C. P. Hawkins, and J. V. Sickie. 2008. Effects of regionalization decisions on an
1514 O/E index for the US national assessment. *Journal of the North American Benthological
1515 Society* **27**:892-905.
- 1516 Zak, D., M. Hupfer, A. Cabezas, G. Jurasinski, J. Audet, A. Kleeberg, R. McInnes, S. M.
1517 Kristiansen, R. J. Petersen, H. Liu, and T. Goldhammer. 2021. Sulphate in freshwater
1518 ecosystems: A review of sources, biogeochemical cycles, ecotoxicological effects and
1519 bioremediation. *Earth-Science Reviews* **212**:103446.

1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538

For Review Only

1539 **Supplementary Information**

1540 **A model-based assessment of anthropogenic disturbance on lotic macroinvertebrate assemblages**

1541
1542 **Darin A. Kopp^{1,2*}, John L. Stoddard², Philip R. Kaufmann^{2,3}, Alan T. Herlihy³, Ryan Hill²,**

1543 **Meredith Brehe^b, Robert Sabo⁶**

1544
1545 **^aOak Ridge Institute for Science and Education Participant c/o US EPA, ORD, Corvallis, Oregon**

1546 ²US EPA, Office of Research and Development, Corvallis, Oregon

1547 ³Department of Fisheries, Wildlife, & Conservation Sciences, Oregon State University,
1548 Corvallis, Oregon

1549 ⁴Current affiliation: Oregon Department of Fish and Wildlife, Salem, Oregon

1550 ⁵Oak Ridge Institute for Science and Education Participant c/o US EPA, ORD, Washington, D.C.

1551 ⁶US EPA, Office of Research and Development, Center for Public Health and Environmental
1552 Assessment, DC, Washington, USA

1554 *Correspondence: kopp.darin@epa.gov

1555
1556
1557
1558
1559
1560
1561
1562 Table S1: Geoclimatic and anthropogenic variables used in random forest models. Variables
1563 were selected based on their hypothesized relationship with total nitrogen, total phosphorus,
1564 chloride, sulfate and substrate diameter. StreamCat data is publicly accessible at
1565 <https://www.epa.gov/national-aquatic-resource-surveys/streamcat-dataset> (Hill et al. 2016)

Variable	Description	Source
BFIWs	The component of streamflow that can be attributed to ground-water discharge	StreamCat; Hill et al. 2016
ClayWs	Mean % clay content of soils in watershed	StreamCat
ElevWs	Mean elevation of watershed (m)	StreamCat
KffactWs	Mean of STATSGO-Kffactor	StreamCat
DepClWs	Mean atmospheric chloride deposition 2018	National Atmospheric Deposition Program
DepPWs	Atmospheric phosphorous deposition 2013	Sabo et al 2023
NWs	Rock-derived nitrogen in watershed	StreamCat
P2O5Ws	Mean % of lithological phosphorous-oxide (P2O5) content in surface or near surface geology	StreamCat
PctAlluvCoastWs	% of watershed area classified with lithology type as alluvium and fine-textured coastal zone sediment	StreamCat
PctHbWetWs	% of watershed area classified as herbaceous wetland land-cover	StreamCat
PermWs	Mean permeability (cm/hour) of soils (STATSGO) within watershed	StreamCat
RunoffWs	Mean runoff (mm) within watershed	StreamCat
SandWs	Mean % sand content of soils (STATSGO) within watershed	StreamCat
StreamPower	The amount of energy the water exerts on the sides and bottom of a stream	$((TPRCP - ET)/1000) * (0.032 * WSArea)^{0.5} * \text{slope}$
SWs	Mean % of lithological sulfur (S) content in surface or near surface geology within watershed	StreamCat
WsAreaSqKm	Area of watershed (square km)	StreamCat
CoalMineDensWs	Density of coal mines sites within watershed (mines/square km)	StreamCat
DepNWs	Mean atmospheric nitrogen deposition in watershed 2018	National Atmospheric Deposition Program
DepSWs	Mean atmospheric sulfur deposition in watershed 2018	National Atmospheric Deposition Program
MineDensWs	Density of coal mines sites within watershed (mines/square km)	StreamCat
N_inputs	Sum of anthropogenic nitrogen inputs: N_Fert_FarmWs, N_Fert_UrbanWs, N_Human_WasteWs, N_Livestock_WasteWs	Sabo et al. 2023
P_inputs	Sum of anthropogenic phosphorus inputs: P_fertilizerWs, P_human_wasteWs, P_livestock_WasteWs, P_nf_fertilizerWs	Sabo et al. 2023
PctCropWs	% of watershed area classified as crop land-use	StreamCat
PctCropWsRp100	% of watershed riparian area classified as crop land-use	StreamCat
PctNatWs	% of watershed with natural vegetation cover	StreamCat
TPRCP	Mean total annual precipitation in watershed	PRISM
RdDensWs	Density of roads (2010 Census Tiger Lines) within watershed (km/square km)	StreamCat
MSAT	Mean summer (July/August) air temperature	PRISM
W1_HAG	Agricultural disturbance adjacent to a stream	NRSA Field Data
NABD_NrmStorWs	Volume all reservoirs per unit area of watershed (cubic meters/square km)	StreamCat

1567 **Table S2:** Variable importance rankings for random forest models and values used to estimate
 1568 physiochemical conditions in the absence of anthropogenic disturbance. The 5 most important
 1569 variables for each model are bold.

Predictor Variable	NTL	PTL	CL	SO4	SUBD	Value Without Disturbance
Geoclimatic						
BFIWs	6	13	5	9	—	—
ClayWs	—	10	7	11	—	—
ElevWs	7	5	8	8	—	—
KffactWs	14	12	15	12	9	—
DepCIWs	—	—	9	—	—	—
DepPWs	—	4	—	—	—	—
NWs	11		12	4	—	—
P2O5Ws	15	6	18	16	—	—
PctAlluvCoastWs	—	—	—	—	5	—
PctHbWetWs	9	7	10	17	—	—
PermWs	13	11	17	13	—	—
RunoffWs	2	2	4	1	2	—
SandWs	—	—	—	—	7	—
StreamPower	—	—	—	—	1	—
SWs	12	16	6	2	—	—
WsAreaSqKm	16	14	13	15	6	—
Anthropogenic						
CoalMineDensWs	—	—	19	7	13	0
DepNWs	5	—	—	—	—	5 kgN/ha/yr*
DepSWs	—	—	14	10	—	1.65 kgS/ha/yr*
MineDensWs	—	—	16	18	11	0
N_inputs	4	—	—	—	—	0
P_inputs	—	9	—	—	—	0
PctCropWs	1	1	3	3	—	0
PctCropWsRp100	—	—	—	—	4	0
PctNatWs	—	—	—	—	3	100
TPRCP	10	15	11	14	—	Average 1940–1950
RdDensWs	3	8	1	6	10	0
MSAT	8	3	2	5	—	Average 1940–1950
W1_HAG	—	—	—	—	8	0
NABD_NrmStorWs	—	—	—	—	12	0

1570 * Estimates for atmospheric nitrogen and sulfur deposition were obtained from Clark et al. 2018.
 1571 Although they used 0.4 kgN/ha and 0.1 kgS/ha no sites had deposition values below this level.
 1572 Instead, we selected the higher values of 3–5 kgN/ha which could be reasonable estimates for
 1573 background deposition values before 1900. For S deposition, other efforts have estimated pre-
 1574 industrial S deposition at 0.32–2.98 kgS/ha (Granat et al. 1976, Fakhraei et al. 2016) and we used
 1575 the used middle number of this interval.

1576 **Table S3.** Regional means of present-day (PD) and hindcasted (HC) values for each environmental gradient. Parentheses are 10th and
 1577 90th quantiles of observed hindcasted values. NTL = Total Nitrogen, PTL = Total Phosphorus, CL = Chloride, SO4 = Sulfate, SUBD =
 1578 Substrate Diameter, RPDI = Riparian Disturbance index, MSAT = Mean Summer Air Temperature, PRCP = Total Annual
 1579 Precipitation.

	NTL (µg/L)		PTL (µg/L)		CL (mg/L)		SO4 (mg/L)		SUB (mm)		RPDI		MSAT (DegC)		TPRCP (mm)	
	PD	HC	PD	HC	PD	HC	PD	HC	PD	HC	PD	HC	PD	HC	PD	HC
CPL	780 (305- 2452.2) 245.58	363 (289- 558) 245.52	82.6 (23.08- 245.52) 77.65	46.74 (25.73- 64.71) 75.51	9.89 (2.93- 7.65) 7.65	4.06 (2.92- 44.55) 16.05	7.49 (1.1- 14.73) 16.05	3.56 (1.93- 6.24) 3.24	0.35 (0.01- 121.32) 11.66	6.17 (1.96- 52.76) 16.64	0.5 (0.03- 1.79) 0.65	0.33 (0.03- 0.33) 0.33 (0- 0.202) 0.33	27.46 (25.54- 28.54) 21.49	27.02 (24.24- 27.85) 20.43	1613.83 (1186.55- 1989.95) 1295.41	1219.48 (1014.36- 1680.13) 1088.38
NAP	431 (187.4- 1309.4) 301.6	263 (214- 75.51) 29.31	23.76 (7.38- 29.31) 17.37	16.73 (0.47- 72.34) 1.97	1.97 (0.3- 5.76) 5.58	5.58 (1.71- 16.05) 3.24	3.24 (1.74- 6.24) 11.66	11.66 (0.08- 121.32) 16.64	16.64 (6.1- 52.76) 16.64	0.65 (0.08- 2.08) 0.65	0.33 (0- 0.33) 0.33 (0- 0.202) 0.33	21.49 (19.01- 23.67) 21.49	20.43 (18.31- 21.78) 20.43	1295.41 (1104.73- 1627.51) 1088.38	1219.48 (1014.36- 1680.13) 1088.38	
NPL	879 (266- 2188) 687.5	456 (202- 361.74) 108.45	85.76 (18.4- 50.42) 76.25	76.25 (33.96- 18.45) 12.09	12.09 (2.35- 18.33) 10.24	10.24 (1.53- 1350.29) 486.24	486.24 (19.03- 653.05) 324.9	324.9 (9.33- 17.94) 0.51	3.44 (0.01- 23.73) 1.27	1.27 (1.13- 2.08) 0.33	0.33 (0.38- 0.33) 0.33	20.52 (18.35- 22.48) 21.19	21.19 (19.1- 23.77) 21.19	437.42 (320.7- 631.93) 360.52	360.52 (282.21- 428.35)	
SAP	557.5 (173.5- 1868) 346.5	265 (169- 128.09) 45.6	29.3 (9.84- 45.6) 26.5	26.5 (14.84- 47.42) 5.89	5.89 (1.04- 3.98) 2.45	2.45 (1.55- 80.35) 7.46	7.46 (1.89- 7.34) 4.3	4.3 (0.35- 206.31) 16.28	16.28 (7.79- 51.83) 15.51	0.76 (0.05- 2.29) 0.76	0.33 (0.05- 0.33) 0.33	24.76 (22.1- 26.72) 24.07	24.07 (21.37- 1809.01) 1475.78	1475.78 (1106.55- 1283.79) 1110.34	1110.34 (916.58- 1283.79)	
SPL	1175 (424.8- 3229.4) 668.4	411 (263.4- 808.04) 150	150 (25.42- 86.67) 57	57 (36.71- 383.92) 30.36	30.36 (4.55- 18.72) 10.27	10.27 (3.97- 905.64) 88.89	88.89 (7.92- 511.06) 66.72	66.72 (9.5- 20.65) 0.35	0.35 (0.02- 12.72) 4.8	1.04 (0.22- 2.35) 1.04	0.33 (0.22- 0.33) 0.33	25.88 (21.93- 29.65) 25.88	26.16 (22.05- 28.9) 26.16	616.17 (394.96- 1060.16) 534.19	534.19 (313.6- 973.8)	
TPL	1816 (659- 8927.6) 558	411 (329.1- 444.72) 167	167 (53.96- 68.09) 50.42	18.9 (7.24- 59.44) 34.4	5.22 (3.43- 8.97) 34.4	34.4 (12.05- 270.07) 12.8	12.8 (5.57- 164.48) 0.35	0.35 (0.01- 15.97) 11.31	0.83 (0.12- 18.96) 0.83	0.33 (0.12- 2.28) 0.33	23.33 (0.12- 25.27) 23.53	23.53 (20.96- 25.41) 1004.22	1004.22 (21.24- 1387.75) 775.73	775.73 (514.38- 922.32)		
UMW	1168 (355- 5250) 418	350 (265- 165.47) 57.4	57.4 (17.1- 56.07) 35.55	35.55 (16.51- 29.93) 9.97	9.97 (0.71- 4.41) 2.56	2.56 (0.34- 41.51) 9.28	9.28 (1.11- 8.29) 3.71	0.35 (0.01- 12.33) 0.35	6.94 (4.59- 28.61) 6.94	0.62 (0- 2.14) 0.62 (0- 0.33)	0.33 (0- 22.56) 0.33 (0- 0.33)	20.84 (18.14- 22.39) 20.98	20.98 (18.42- 1393.8) 960.16	960.16 (788.56- 877.31) 793.53	793.53 (622.66- 877.31)	
WMT	133 (50- 386.8) 216	133 (65- 82.73) 21.84	21.84 (7.09- 58.72) 21.14	21.14 (8.61- 9.18) 0.92	0.92 (0.12- 2.53) 0.52	0.52 (0.15- 26.75) 2.97	2.97 (0.47- 19.97) 2.87	2.87 (0.93- 202.46) 38.31	38.31 (0.98- 82.76) 34.89	0.48 (0- 1.72) 0.48 (0- 1.72)	0.33 (0- 22.29) 0.33 (0- 0.33)	18.08 (13.98- 21.33) 17	17 (12.61- 21.33) 669.31	669.31 (349.23- 1707.15) 701.68	701.68 (387.1- 1812.51)	
XER	340 (110- 1258.8) 396.4	218 (133- 258.99) 53.87	53.87 (14.35- 90.21) 45.76	45.76 (15.65- 143.8) 5.35	5.35 (0.8- 14.32) 2.9	2.9 (0.73- 403.9) 27.8	27.8 (1.6- 186.07) 22.32	22.32 (2.54- 57.52) 2.02	2.02 (3.53- 45.29) 12.71	1.16 (0.12- 45.29) 1.16	0.33 (0.12- 2.61) 0.33	22.76 (18.79- 30.29) 21.36	21.36 (17.45- 27.99) 287.58	287.58 (169.27- 503.23) 336.66	336.66 (220.95- 694.06)	

For Review Only

1582 Tables

1583 Table 1: Sites, genera, and environmental variables included in the analysis. Revisit sites are locations that were revisited during the
 1584 survey to assess within year variability. Values for the environmental variables are median values for each region. Range is given in
 1585 parentheses. NTL = Total Nitrogen, PTL = Total Phosphorous, CL = Chloride, SO4 = Sulfate, RPDI = Riparian Disturbance Index,
 1586 SUBD = Substrate Diameter, TPRCP = Total Precipitation and MSAT = Mean Summer Air Temperature. CPL = Coastal Plains,
 1587 NAP = Northern Appalachians, NPL = Northern Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL = Temperate
 1588 Plains, UMW = Upper Midwest; WMT = Western Mountains; XER = Xeric.

1589

Region	Site (#)	Genera (#)	Revisit Sites (#)	NTL (ug/L)	PTL (ug/L)	CL (mg/L)	SO4 (mg/L)	RPDI	SUBD (mm)	TPRCP (mm)	MSAT (°C)
CPL	226	71	36	786 (91-7713)	84.81 (5.2-3922.25)	10.06 (0.52-4797.9)	7.83 (0.11-1900.89)	0.5 (0-5.39)	0.35 (0.01-341.65)	1613.83 (388.66-2397.1)	27.46 (23.75-31.68)
NAP	228	127	29	431 (81-6413)	23.76 (3.12-587.2)	16.73 (0.07-668.16)	5.58 (0.06-467.93)	0.65 (0-5.24)	11.66 (0.01-864.9)	1295.41 (888.91-1883.17)	21.49 (17.24-24.72)
NPL	152	76	7	881.5 (71-15675)	85.61 (3.53-9248.31)	12.09 (0.11-1251.58)	486.24 (3.5-4079.78)	1.27 (0-4.49)	0.51 (0.01-560.5)	437.42 (157.84-1055.11)	20.52 (15.62-23.46)
SAP	266	111	32	557.5 (36-18700)	29.3 (3.79-4050)	5.89 (0.36-197.39)	7.46 (0.58-397.69)	0.76 (0-4.56)	16.28 (0.01-5656.85)	1475.78 (889.64-2451.81)	24.76 (19.19-27.73)
SPL	174	59	5	1165 (145-21175)	147.12 (5.21-4351.7)	30.36 (0.43-5220)	88.89 (1.96-3716.6)	1.04 (0-5.88)	0.35 (0.01-5656.85)	616.17 (240.1-1404.61)	25.88 (11.96-32.75)
TPL	223	74	29	1806 (236-16219)	165.08 (11.45-1066.87)	18.9 (1.44-736.62)	34.4 (5.79-1386.77)	0.83 (0-5.47)	0.35 (0.01-5656.85)	1004.22 (376.7-1801.08)	23.33 (18.28-27.09)
UMW	201	104	10	1168 (195-17675)	57.4 (8.16-659)	9.97 (0.01-306.69)	9.28 (0.04-160.4)	0.62 (0-6.43)	0.35 (0.01-812.79)	960.16 (496.44-1718.2)	20.84 (17.05-23.43)
WMT	225	94	19	133 (22-4719)	21.84 (2.71-569.8)	0.92 (0.04-521.64)	2.97 (0.07-1682.05)	0.48 (0-3.95)	38.14 (0.01-1288.61)	669.31 (175.89-3946.33)	18.08 (11.71-29.4)
XER	196	75	16	344.5 (46-8000)	54.62 (4.29-4667.41)	5.35 (0.1-1867.57)	27.8 (0.02-3286.24)	1.16 (0-4.36)	2.02 (0.01-368.11)	287.58 (100.46-1176.56)	22.76 (14.69-36.11)

1591 Table 2: Random Forest model performance metrics for testing and out-of-bag training datasets. NTL = Total
1592 Phosphorus, CL = Chloride, SO₄ = Sulfate, SUBD = Substrate Diameter. RMSE = Root mean squared error of random forest models
1593 fitted with ln(x + 1) (NTL, PTL, and CL) or ln(x) (SO₄) or Log10(SUBD) transformations.

1594

Variable	R ² _{train}	RMSE _{train}	R ² _{test}	RMSE _{test}
ln(NTL+1) (ug/L)	0.73	0.61	0.72	0.60
ln(PTL+1) (ug/L)	0.60	0.78	0.69	0.70
ln(CL+1) (mg/L)	0.73	0.75	0.68	0.85
ln(SO ₄) (mg/L)	0.77	0.96	0.78	0.97
log ₁₀ (SUBD) (mm)	0.46	0.98	0.51	1.00

1595

1596

1597

1598

1599

1600

1601

1602

1603

1604

1605

1606

1607

1608

1609 Table 3. Regression coefficients between predicted and observed richness and compositional similarity. Occurrence probabilities
 1610 thresholds (Pr) were used to exclude genera with low predicted occurrence probabilities. We considered $R^2 \geq 0.2$, $-1.5 \leq \text{intercept} \leq$
 1611 1.5 , and $0.85 \leq \text{slope} \leq 1.15$ are indicative of adequate model performance. Values in parentheses are the 5th and 95th percentile of
 1612 estimates from the 3000 posterior samples. Compositional similarity was measured using a probabilistic adaptation of Jaccard index
 1613 measured for each site. Parentheses are the 5th and 95th percentile of values for all sites.

	Pr	Genus Richness				Composition	
		Intercept	Slope	R ²	Revisit Sites within Posterior Distribution (%)	Mean Jaccard Similarity	Mean Jaccard Similarity at revisit sites
CPL	0.05	0.37 (-0.37 - 1.02)	1.00 (0.96 - 1.05)	0.85 (0.82 - 0.87)	97	0.52 (0.26 - 0.7)	0.68 (0.24-0.95)
NAP	0.05	-0.45 (-2.02 - 1.02)	1.03 (0.99 - 1.08)	0.74 (0.71 - 0.78)	90	0.52 (0.31 - 0.64)	0.35 (0.17-0.65)
NPL	0.05	-0.07 (-1.21 - 1.04)	1.02 (0.96 - 1.09)	0.76 (0.71 - 0.81)	100	0.58 (0.34 - 0.77)	0.39 (0.12-0.71)
SAP	0.05	0.40 (-0.66 - 1.41)	1.00 (0.97 - 1.04)	0.85 (0.82 - 0.87)	97	0.52 (0.26 - 0.67)	0.56 (0.21-0.87)
SPL	0.10	0.48 (-0.45 - 1.29)	1.02 (0.97 - 1.09)	0.79 (0.75 - 0.83)	100	0.52 (0.25 - 0.72)	0.61 (0.52-0.98)
TPL	0.05	0.06 (-0.91 - 0.94)	1.01 (0.97 - 1.06)	0.82 (0.79 - 0.84)	100	0.55 (0.31 - 0.77)	0.59 (0.18-1)
UMW	0.05	-0.74 (-2.52 - 0.99)	1.04 (0.98 - 1.10)	0.73 (0.66 - 0.78)	100	0.55 (0.34 - 0.69)	0.53 (0.27-0.9)
WMT	0.10	0.24 (-1.24 - 1.59)	1.04 (0.99 - 1.10)	0.74 (0.70 - 0.78)	100	0.60 (0.36 - 0.75)	0.45 (0.16-0.79)
XER	0.05	-0.12 (-1.11 - 0.74)	1.03 (0.98 - 1.08)	0.83 (0.80 - 0.86)	94	0.50 (0.23 - 0.71)	0.48 (0.07-0.83)

1618 Table 4: Mean proportion of sites in each taxonomic group that was identified as an increaser or decreaser. Increasers (I) are genera
 1619 that have a significantly higher probability of occurrence under present day conditions compared to hindcasted conditions; decreasers
 1620 (D) are genera that have a significantly lower probability of occurrence under present-day conditions compared to hindcasted
 1621 conditions. The values in the table are the mean proportion of sites for genera within the major taxonomic group. CPL = Coastal
 1622 Plains, NAP = Northern Appalachians, NPL = Northern Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL =
 1623 Temperate Plains, UMW = Upper Midwest; WMT = Western Mountains; XER = xeric. EPT = Ephemeroptera, Plecoptera and
 1624 Trichoptera.

	<u>CPL</u>		<u>NAP</u>		<u>NPL</u>		<u>SAP</u>		<u>SPL</u>		<u>TPL</u>		<u>UMW</u>		<u>WMT</u>		<u>XER</u>	
	<u>D</u>	<u>I</u>																
<u>Insects</u>	<u>0.16</u>	<u>0.08</u>	<u>0.08</u>	<u>0.05</u>	<u>0.1</u>	<u>0.17</u>	<u>0.1</u>	<u>0.11</u>	<u>0.09</u>	<u>0.12</u>	<u>0.16</u>	<u>0.11</u>	<u>0.11</u>	<u>0.12</u>	<u>0.03</u>	<u>0.06</u>	<u>0.09</u>	<u>0.05</u>
<u>EPT</u>	<u>0.24</u>	<u>0.01</u>	<u>0.07</u>	<u>0.04</u>	<u>0.17</u>	<u>0.11</u>	<u>0.13</u>	<u>0.08</u>	<u>0.09</u>	<u>0.04</u>	<u>0.23</u>	<u>0.04</u>	<u>0.14</u>	<u>0.03</u>	<u>0.04</u>	<u>0.06</u>	<u>0.1</u>	<u>0.03</u>
<u>CHIRONOMIDAE</u>	<u>0.11</u>	<u>0.09</u>	<u>0.09</u>	<u>0.05</u>	<u>0.04</u>	<u>0.18</u>	<u>0.04</u>	<u>0.12</u>	<u>0.02</u>	<u>0.17</u>	<u>0.12</u>	<u>0.14</u>	<u>0.06</u>	<u>0.16</u>	<u>0.02</u>	<u>0.07</u>	<u>0.06</u>	<u>0.1</u>
<u>Other Insects</u>	<u>0.21</u>	<u>0.07</u>	<u>0.08</u>	<u>0.08</u>	<u>0.07</u>	<u>0.21</u>	<u>0.13</u>	<u>0.12</u>	<u>0.23</u>	<u>0.02</u>	<u>0.18</u>	<u>0.11</u>	<u>0.09</u>	<u>0.04</u>	<u>0.02</u>	<u>0.03</u>	<u>0.11</u>	<u>0.01</u>
<u>Non-Insects</u>	<u>0.14</u>	<u>0.16</u>	<u>0.06</u>	<u>0.18</u>	<u>0.06</u>	<u>0.19</u>	<u>0.11</u>	<u>0.13</u>	<u>0.06</u>	<u>0.13</u>	<u>0.15</u>	<u>0.15</u>	<u>0.08</u>	<u>0.14</u>	<u>0.03</u>	<u>0.08</u>	<u>0.04</u>	<u>0.14</u>
<u>ARTHROPODA</u>	<u>0.21</u>	<u>0.11</u>	<u>0.07</u>	<u>0.16</u>	<u>0.11</u>	<u>0.02</u>	<u>0.16</u>	<u>0.09</u>	<u>0.05</u>	<u>0.06</u>	<u>0.23</u>	<u>0.35</u>	<u>0.1</u>	<u>0.04</u>	<u>0.04</u>	<u>0.03</u>	<u>0.05</u>	<u>0.04</u>
<u>MOLLUSCA</u>	<u>0.08</u>	<u>0.11</u>	<u>0.03</u>	<u>0.28</u>	<u>0.02</u>	<u>0.3</u>	<u>0.1</u>	<u>0.15</u>	<u>0.09</u>	<u>0.06</u>	<u>0.21</u>	<u>0.05</u>	<u>0.14</u>	<u>0.07</u>	<u>0.01</u>	<u>0.06</u>	<u>0.01</u>	<u>0.26</u>
<u>Other Non-Insects</u>	<u>0.15</u>	<u>0.22</u>	--	<u>0.14</u>	<u>0.05</u>	<u>0.13</u>	<u>0.04</u>	<u>0.16</u>	<u>0.02</u>	<u>0.19</u>	<u>0.05</u>	<u>0.15</u>	<u>0.01</u>	<u>0.25</u>	<u>0.04</u>	<u>0.14</u>	<u>0.04</u>	<u>0.14</u>

1626 Figure Captions

1627 Figure 1: Ecoregions and survey locations for the National Rivers and Streams Assessment
1628 2018-2019 survey. CPL = Coastal Plains, NAP = Northern Appalachians, NPL = Northern
1629 Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL = Temperate Plains, UMW =
1630 Upper Midwest; WMT = Western Mountains; XER = Xeric

1631

1632 Figure 2: Conceptual diagram of posterior distributions generated from bayesian inference. The
1633 grey is a posterior distribution of genus richness generated after removing anthropogenic
1634 disturbances (Hindcasted) and the white is a posterior distribution generated from present-day
1635 conditions. The dotted line is the expected value (i.e. mean) of the present day posterior distribution
1636 and dark shading represents the lower 10% or the upper 90% of the HC posterior distribution. At
1637 site A, the PD mean is within the <10% of the HC distribution and indicates that PD richness is
1638 likely lower than HC richness. Alternatively, at site B, PD richness is >90% of the HC distribution
1639 and indicates that PD richness is likely higher. At site C, the two posterior distributions are similar
1640 such that there is likely no difference between PD and HC.

1641

1642 Figure 3: Partial dependence plots showing the effects of anthropogenic variables on in-stream
1643 physicochemical factors. For visualization, each anthropogenic factor was rescaled between 0-1
1644 and labeled low, medium, and high. CoalMineDen = coal mine density, MineDen = gravel mine
1645 density, MSAT = mean summer air temperature, N_dep = atmospheric nitrogen deposition,
1646 N_input = anthropogenic nitrogen inputs, P_input = anthropogenic phosphorous inputs, PctCrop
1647 = percent crop in the watershed, PctCropRP = percent crop in the riparian area, PctNatRP =
1648 percent natural vegetation in riparian area, RdDen = road density, S_dep = atmospheric sulfur
1649 deposition, TPRCP = total precipitation, and W1_HAG = agricultural disturbance adjacent to
1650 stream reach. NTL = Total Nitrogen (ug/L), PTL = Total Phosphorus (ug/L), CL = Chloride
1651 (mg/L), SO4 = Sulfate (mg/L), SUBD = Substrate Diameter Log10(mm).

1652

1653 Figure 4: Observed versus hindcasted values for each environmental gradient. Points are regional
1654 means and vertical and horizontal bars represent the 10th and 90th quantiles of observed values or
1655 hindcasted values within each region, respectively. For RPDI, all hindcast values were < 0.33
1656 and the ecoregions were plotted separately for visualization. The dashed line is the 1:1
1657 relationship for all plots except for RPDI where it represents the 0.33 threshold applied to all
1658 ecoregions. NTL = Total Nitrogen (ug/L), PTL = Total Phosphorus (ug/L), CL = Chloride (mg/L),
1659 SO4 = Sulfate (mg/L), SUBD = Substrate Diameter (mm), RPDI = Riparian Disturbance Index,
1660 MSAT = mean summer air temperature (°C) and TPRCP = total precipitation (mm). CPL =
1661 Coastal Plains, NAP = Northern Appalachians, NPL = Northern Plains, SAP = Southern
1662 Appalachians, SPL = Southern Plains, TPL = Temperate Plains, UMW = Upper Midwest; WMT
1663 = Western Mountains; XER = Xeric. See Appendix S1: Table S4 for observed and hindcasted
1664 mean values and quantiles for each region.

1665

1666 Figure 5: A) Proportion of sites where hindcasted abiotic conditions were >2 standard deviations
1667 from the observed value for Total Nitrogen (NTL), Total Phosphorus (PTL), Chloride (CL),
1668 Sulfate (SO₄), Substrate Diameter (SUBD), mean summer air temperature (MSAT), and Total
1669 Precipitation (TPRCP) and > 0.33 for Riparian Disturbance Index (RPDI). B) The number of
1670 environmental variables affected by human disturbance. Locations where the effects were not
1671 detected are plotted separately. CPL = Coastal Plains, NAP = Northern Appalachians, NPL =
1672 Northern Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL = Temperate
1673 Plains, UMW = Upper Midwest; WMT = Western Mountains; XER = Xeric.

1674

1675 Figure 6: Difference in macroinvertebrate genus richness after removing anthropogenic
1676 disturbance from each category of physiochemical variables (A) and all variables simultaneously
1677 (B). The points in each panel are sites that had a change in genus richness with >75% support
1678 after removing disturbance.

1679

1680 Figure 7: Percentage of the population of streams where genus richness or composition could
1681 change given a hindcasted physiochemical environment. Shaded bars indicate that the probability
1682 of mean present-day richness differs from hindcast with > 0.90 (Black) or 0.75-0.90 (Gray)
1683 support. Hatched bars indicate that Jaccard similarity was < 0.9 but support for a difference in
1684 richness were < 0.75. White bars indicate the proportion of sites that had < 0.75 support for
1685 change in richness and >0.9 compositional similarity. Yellow bars indicate the proportion of sites
1686 that could not be assessed because of insufficient data or potential extrapolation from predicting
1687 hindcast physiochemical conditions. Regions are arranged according to richness. NPL =
1688 Northern Plains, TPL = Temperate Plains, SPL = Southern Plains, XER = Xeric, SAP =
1689 Southern Appalachians, CPL = Coastal Plains, NAP = Northern Appalachians, UMW = Upper
1690 Midwest and WMT = Western Mountains.

1691