1 **Title:** Phylogenetic diversity efficiently and accurately prioritizes the conservation of aquatic
2 macroinvertebrate communities

3 **Running head:** Phylogenetic assessment of aquatic communities

4 **Authors:** Kathryn M. Weglarz^{1*}, W. Carl Saunders², Andrew Van Wagenen², & William D.
5 Pearse¹

6 ¹ Department of Biology & Ecology Center, Utah State University, 5305 Old Main Hill, Logan, UT
7 84322

8 ² PACFISH/INFISH Biological Observation Monitoring Effectiveness Program, USDA Forest Ser-
9 vice, Intermountain Region, Region 4, 860 North 1200 East, Logan, UT 84321

10 *kathryn.weglarz@usu.edu

11 **Keywords:** Phylogenetic diversity, biological assessment, aquatic, macroinvertebrates, monitoring,
12 metric

Abstract

1. Land degradation is a leading cause of biodiversity loss yet its consequences on freshwater ecosystems are poorly understood, exacerbating difficulties with assessing ecosystem quality and the effectiveness of restoration practices.
2. Many monitoring programs rely on macroinvertebrates to assess the biotic effects of degradation and/or restoration and management actions on freshwater ecosystems. The ratio of Observed (O) to Expected (E) macroinvertebrate taxa at a given site—O/E—is often used for this purpose, despite the amount of modeling and data required to generate expectations and difficulties quantitatively assessing the degree of degradation at a site.
3. Despite widespread use in academic biology, phylogenetic diversity is rarely applied in management, regardless of empirical correlations between phylogenetic diversity and management targets such as ecosystem structure and function.
4. We use macroinvertebrate data from 1,400 watersheds to evaluate the potential for phylogenetic metrics to inform evaluations of management practices. These data have been collected since 1998, and have been used to determine the effectiveness of conservation management for the maintenance and restoration of riparian and aquatic systems.
5. Phylogenetic diversity detected degradation as effectively as O/E, despite not having baseline ‘expectation’ data. Site disturbance, road density, and broader environmental drivers such as mean annual temperature strongly predicted site phylogenetic diversity, providing concrete management objectives to increase site health.
6. *Synthesis and applications.* Management efforts targeted solely at taxonomic metrics, such as O/E, have been successfully used to manage sites. We show here that phylogenetic diversity metrics can support such efforts by providing additional information about the kind of species at sites. Given the ease with which such approaches can be applied, we call on others to use them to supplement existing prioritization schemes.

1 Introduction

Declines in biodiversity are associated with subsequent loss of ecosystem services (Butchart et al., 2010; Cardinale et al., 2012). This is a particular concern in freshwater ecosystems which are highly vulnerable (Millennium Ecosystem Assessment, 2005). These ecosystems support food web processes and nutrient cycling, purify and supply high quality freshwater resources, regulate sediment and nutrient transport across the landscape, and provide an array of cultural services (Macadam & Stockan, 2015; Vörösmarty et al., 2010). Since the late 1800s, when the link between human health and water quality was formally recognized by the scientific community, freshwater systems have been monitored for conservation and restoration purposes (Bonada, Prat, Resh, & Statzner, 2006). While land degradation is recognized as a leading cause of biodiversity loss (Newbold et al., 2015), its consequences on freshwater ecosystems remain poorly understood (IPBES, 2018). This gap persists in spite of the well-recognized link between land and water management (Bossio, Geheb, & Critchley, 2010), making it difficult to determine the effectiveness of restoration practices (Feld et al., 2011; Melland, Fenton, & Jordan, 2018). To assess the impact of these practices on the biotic community, many monitoring programs survey aquatic macroinvertebrates because these organisms are abundant, provide anywhere from 20-100% of the energy budget to consumers, and are sensitive to changes in water chemistry in predictable ways (Bonada et al., 2006; Macadam & Stockan, 2015). Beyond serving as a biomonitoring tool, macroinvertebrates are integral to decomposition and nutrient cycling and support \$31.4 billion in recreational fishing in the US alone (Prather et al., 2013). Unfortunately, the impact of management efforts on freshwater biodiversity, particularly macroinvertebrates and fish, can be difficult to detect owing to the time required for these organisms to return to improved habitats (Meals, Dressing, & Davenport, 2010).

Many monitoring programs rely on macroinvertebrates to assess the biotic effects of degradation because they are relatively easy to collect and identify, and track disruptions more closely than other taxonomic groups (Bonada et al., 2006; Cairns & Pratt, 1993; Friberg et al., 2011). Due to their ubiquity in monitoring, a number metrics have been developed to evaluate the status of macroinvertebrate assemblages (Balloch, Davies, & Jones, 1976; Bonada et al., 2006; Cairns & Pratt, 1993; Hellawell, 2012; Resh & Rosenberg, 1993). Often, these metrics measure ecological structure

and are similar to, or based upon, measures of species richness (Yates, Brua, Culp, Chambers, & Wassenaar, 2014). One of the more commonly adopted families of these metrics compares the number of Observed taxa (O) to a statistical Expectation (E) of taxa at a given site (O/E; Hawkins, 2006; Hawkins, Norris, Hogue, & Feminella, 2000; Wright, 1995). This metric became widely applied in management with the development and adoption of the River Invertebrate Prediction and Classification System (RIVPACS) (Wright, Sutcliffe, & Furse, 2000; Wright, 1995), which has since been modified for regions outside of the UK (Hawkins et al., 2000; Moya et al., 2011; Smith, Kay, Edward, Papas, et al., 1999). Such O/E approaches are comparable across systems and relatively easy for managers to interpret (Hawkins, 2006; Wright et al., 2000; Wright, 1995), but, like all approaches, O/E metrics have potential caveats. They are less sensitive to degradation in heterogeneous environments (Hargett, ZumBerge, Hawkins, & Olson, 2007), and uncertainty associated with expectations is often not reflected in the final O/E metric (but see de Zwart, Dyer, Posthuma, & Hawkins, 2006). Further, standards for what constitutes a degraded O/E are often subjective, and statistical deviation from this baseline can be difficult to determine (Hämäläinen, Aroviita, Jyväsjärvi, & Kärkkäinen, 2018). The problem is even more acute given variation (and uncertainty; see above) in expectations of site richness: sites may have the same O/E value even when they have drastically different values of E. For instance, a site with an $\frac{O}{E}$ of $\frac{4}{5}$ and another with an $\frac{O}{E}$ of $\frac{80}{100}$ represent the same O/E value of 0.8, but it is unclear whether this value conveys the same information for both sites. Perhaps most critically, however, O/E does not consider the *kinds* of species that are present or absent. The functional diversity of macroinvertebrates reflects an ecosystem's health and productivity (Schmera, Heino, Podani, Erős, & Dolédec, 2017), and it would be useful, as others have suggested (*e.g.*, Friberg et al., 2011; Yates et al., 2014), for metrics to reflect this.

Even though integrating functional and structural metrics should improve detection of environmental degradation (Yates et al., 2014), macroinvertebrate modeling has been slow to incorporate phylogenetic diversity, which can serve as a proxy for functional diversity (Cavender-Bares, Kozak, Fine, & Kembel, 2009; Faith, 1992; Tucker et al., 2017; Webb, Ackerly, McPeck, & Donoghue, 2002). This is surprising, since much of modern eco-phylogenetic methods were developed from

taxonomic metrics developed for macroinvertebrates (reviewed in Pearse, Purvis, Cavender-Bares, and Helmus, 2014; see Izsáki and Papp, 1995; Wright, 1995). Incorporating the evolutionary history of species (phylogeny) into macroinvertebrate monitoring provides the opportunity to incorporate information about the kinds of species within an ecosystem. Under the assumption that more distantly-related species have fewer functional traits in common (but see Mayfield & Levine, 2010), phylogenetic diversity metrics can reveal the structure (Cadotte, Hamilton, & Murray, 2009), stability (Cadotte, Dinnage, & Tilman, 2012; Craven et al., 2018), and primary productivity (Cadotte, Cardinale, & Oakley, 2008; Flynn, Mirotchnick, Jain, Palmer, & Naeem, 2011), of ecosystems while highlighting when certain species loss or gain will have an oversized impact on ecosystem function (Lessard, Fordyce, Gotelli, & Sanders, 2009; Pearse, Chase, et al., 2015; Strauss, Webb, & Salamin, 2006). Yet phylogenetic diversity or evolutionary distinctiveness are rarely incorporated into applied conservation, restoration, and management programs (Cadotte & Tucker, 2018; Díaz et al., 2013; Pearse, Chase, et al., 2015; Tucker et al., 2017), with the notable exception of the EDGE of Existence program (Isaac & Pearse, 2018; Isaac, Turvey, Collen, Waterman, & Baillie, 2007).

The goal of this study is to outline the potential of phylogenetic metrics for macroinvertebrate conservation and monitoring programs. To accomplish this we use a well-established, successful monitoring program as a case study: the PACFISH/INFISH Biological Opinion Effectiveness Monitoring Program (PIBO). PIBO's goal is to determine the effectiveness of aquatic conservation strategies in riparian and aquatic systems in the Northwestern US, and has sampled stream reaches in over 1,400 watersheds every five years since the late 1990s (see Fig. 1; Henderson, Archer, Bouwes, Coles-Ritchie, & Kershner, 2005). The program's macroinvertebrate assemblage data provides a direct link between sites' physical characteristics and the broader biotic community, as macroinvertebrates are prey for the imperiled fishes PIBO was designed to assess. Yet most analyses of the PIBO dataset have focused on physical habitat measures (*e.g.*, Al-Chokhachy et al., 2016; Al-Chokhachy, Roper, Archer, & Miller, 2011; Kershner et al., 2004), and initial attempts to use traditional metrics such as O/E provided insufficient information to evaluate management actions (Irvine et al., 2015). Here, we apply phylogenetic diversity metrics to the macroinvertebrate data within the PIBO dataset to demonstrate their usefulness to biomonitoring programs. We focus on

the analysis of three phylogenetic diversity metrics (Faith’s PD, SES_{MPD} , and SES_{MNTD}) as a case study in the potential of phylogenetic diversity metrics to generate actionable management insights.

2 Methods

2.1 Data Collection

The PIBO program collects data from within the Interior Columbia River and Upper Missouri River basins at reach-scale sites (160-400m stream lengths). Sites designated as ‘reference’ (minimally managed) are primarily located in watersheds that have little history of, or no obvious, mining, recent grazing (within 30 years), timber harvest ($< 5\%$), and have minimal road density within the watershed ($< 0.5 \text{ km/km}^2$). Macroinvertebrates are sampled from eight fast-water habitats per site and combined into a composite sample using the protocol recommended by the Center for Monitoring and Assessment of Freshwater Ecosystems at Utah State University (Hawkins et al., 2000). Taxa are identified by the BLM/USU National Aquatic Monitoring Center in Logan (Utah). Physical measures of the environment, such as the frequency of large wood, total dissolved solids, bankfull width, and pool frequency, are sampled concurrently across the PIBO sites. For a complete description of these variables and field methods used, see Henderson et al. (2005). Here we analyze PIBO data collected between 2001 and 2017 across 1667 sites, of which 1477 (88%) were sampled more than once.

2.2 Measures of Site Disturbance

All following analyses were conducted in R (3.5.1; R Core Team, 2018), data and code for analyses are available in supplementary materials. To build a macroinvertebrate phylogeny, we searched the TimeTree of Life (Hedges, Marin, Suleski, Paymer, & Kumar, 2015) for species and their congeners, and then added missing species into this phylogeny using dating information from congeners using the ‘bind-replace’ algorithm (*congeneric.merge* in *pez*; Pearse, Cadotte, et al., 2015). The resulting

phylogeny contained 164 of the 440 insect taxa in the PIBO data; this was the best coverage we could achieve given many species were identified only to genus (260; and so could not be included in a species-level phylogenetic analysis) and the paucity of genetic data available. The strong performance of our phylogenetic metrics, despite this lack of data, makes our demonstration of the power of phylogenetic metrics, if anything, conservative. However, we note that phylogenetic diversity analyses are resilient to randomly missing species (Isaac & Pearse, 2018). Using this phylogeny, we calculated Faith's PD (Faith, 1992), SES_{MPD} (Kembel, 2009; Webb et al., 2002), and SES_{MNTD} using *pez* (Kembel et al., 2010; Pearse, Cadotte, et al., 2015) under a species-labels null randomization. To compare the relative power of O/E and phylogenetic metrics, we also analyzed the O/E metric values that PIBO calculates for management purposes. PIBO calculates O/E by employing a RIVPACS-type predictive model and comparing the number of taxa expected at similar high quality sites to the number observed at a site (for details see Irvine et al., 2015). This resulted in metrics for 5033 site-year combinations (1477 of the 1667 sites were surveyed more than once) within the PIBO dataset.

2.3 Drivers of biotic degradation

To explain potential variation in phylogenetic and O/E metrics, we utilized a range of covariates that describe variation in geoclimatic and anthropogenic drivers across the study area. Table S1 outlines the explanatory variables used in our analysis. We divided these variables into 'actionable' variables that management practices may affect, and 'non-actionable' which management is unlikely to be able to control. We included the PIBO program's site designation as 'managed' or 'reference'. To earn the 'reference' designation, sites must have no grazing in the watershed (in the last 30 years), <5% of the watershed harvested for timber, no mining (in the last 30 years), and a <0.5 km/km² road density in the watershed. We also included the PIBO program's condition index of habitat integrity. This condition index is a numeric score ranging from 0 (worst) to 100 (best), calculated by summing independent index values of six variables: residual pool depth, percent pools, diameter of the 50th percentile streambed particle, percent pool tail fines <6mm, large wood frequency, and average bank angle (see Al-Chokhachy, Roper, and Archer (2010) for all details).

With over 40 potential explanatory variables in our dataset, we identified potential drivers of degradation using a two-step approach. First, we used lasso regression (Hastie & Efron, 2013; Tibshirani, 1996) to eliminate explanatory variables with minimal sway on response variables. Lasso regression is a machine learning algorithm designed to reduce a large number of potential explanatory variables to a suite of those that drive a response variable (Hastie & Efron, 2013). With the set of 19 explanatory variables identified by the lasso regression as potentially important for at least one of our four response variables (Faith’s PD, SES_{MPD} , SES_{MNTD} , and O/E), we then used information theoretic criteria to identify the most important predictors of our response variables (Bartoń, 2018). Information theoretic criteria (Burnham & Anderson, 2004) allow for uncertainty in both model specification (which variables are important) and parameters (coefficient estimates) to filter through into model predictions. This allows us to circumvent traditional problems with significance thresholds and their arbitrary decision criteria when dealing with datasets of this size and complexity. We fit mixed effects models using site identity as a random covariate to control for repeated measures from sites through time (Bates, Mächler, Bolker, & Walker, 2015). This ensures that our estimates are not biased by changes in sampling through time and space in the data. Finally, we ensured all explanatory variables were Z-transformed prior to analysis, making each variable’s coefficient a measure of relative importance (Gelman, Carlin, Stern, & Rubin, 2014; Grueber, Nakagawa, Laws, & Jamieson, 2011). This makes variable coefficients directly comparable, such that a variable with a coefficient twice that of another variable is twice as important in driving a response variable, and so allows us to generate more actionable insights.

2.4 Identifying sites of concern

To generate site-specific recommendations, we used the random effects of our mixed effect models (Bates et al., 2015). In a mixed effects framework, treating each site identity as a ‘random’ variable controls for uneven sampling of sites while quantifying the impacts of the other explanatory variables. It also generates site-specific estimates for identifying sites with unusually high or low values of our diversity indices (PD, SES_{MPD} , SES_{MNTD} , and O/E) within the context of our model. This method is appropriate for making recommendations within a dataset, but these site-level predic-

tions should not be used to generalize to other nearby sites or to different datasets (as is the case for related random-effects approaches; Hadfield, Wilson, Garant, Sheldon, & Kruuk, 2010; Robinson, 1991). Spatial plotting was conducted using the R packages ‘*rgdal*’, ‘*rgeos*’, ‘*raster*’, and ‘*ggmap*’ (Bivand, Keitt, & Rowlingson, 2018; Bivand & Rundel, 2018; Hijmans, 2018; Kahle & Wickham, 2013).

3 Results

The stream-reach condition index that PIBO has been using to quantify management impacts at sampled stream reaches is significantly negatively correlated with SES_{MNTD} and SES_{MPD} (the SES_{MNTD} correlation is strongest; Fig. 2). Thus, sites with more intact aquatic habitat tend to contain macroinvertebrate communities with more close relatives (phylogenetic clustering). Disturbed sites (*i.e.*, low index scores), by contrast, are more phylogenetically overdispersed, which is consistent with the pattern that may result from invasion by distantly-related (novel) species at these sites (not explored here, but see Proches, Wilson, Richardson, & Rejmánek, 2008). Faith’s PD is the only metric we examined that does not incorporate a statistical expectation of diversity, and it is the only metric not to correlate with the condition index (Fig. 2) and for which ‘reference’ sites differed from ‘managed’ sites (Fig. 3).

Many of the important predictors of O/E, SES_{MPD} , and SES_{MNTD} were the same, while the drivers of PD differed (Fig. 3). Some of the important drivers of SES_{MPD} and SES_{MNTD} are not practical for management interventions, such as stream gradient and latitude. There were, however, several potentially actionable factors related to stream geomorphology (average bankfull width, pool tail fines, and number of pools) and road density (in the reach itself and its surroundings). SES_{MPD} and SES_{MNTD} differed in the size of pool tail fine sediment that affected them, with SES_{MPD} more affected by larger particles ($< 6\text{mm}$) than SES_{MNTD} ($< 2\text{mm}$).

Using our mixed effects framework, we were able to identify sites with unusually high or low (phylogenetic) metric values (Fig. 4). Notably, geographic hot-spots and cold-spots are visible for all variables, which is likely related to the non-actionable diversity drivers we identified, such as

226 latitude.

227 4 Discussion

228 Assessing and tracking the condition of aquatic and riparian resources on federal lands necessitates
 229 tracking changes in sites' abiotic and biotic conditions to ensure management is effective at achieving
 230 its objectives. Many previous analyses of the PIBO dataset use physical habitat measures, compared
 231 between reference and managed sites, to assess stream health (e.g. Al-Chokhachy et al., 2016; Al-
 232 Chokhachy et al., 2011; Kershner et al., 2004). This work identified important drivers of habitat
 233 degradation, but did not quantify the biotic response to such disturbances. Efforts to assess biotic
 234 response have primarily focused on the vegetation component of the PIBO dataset, examining the
 235 impact of grazing on plants (Coles-Ritchie, Roberts, Kershner, & Henderson, 2007), of plants on
 236 stream structure (Hough-Snee, Roper, Wheaton, & Lokteff, 2015; Roper, Jarvis, & Kershner, 2007),
 237 and the physical site factors that predict plant invasion (Al-Chokhachy, Ray, Roper, & Archer, 2013;
 238 Menuz & Kettenring, 2013). These studies are a solid foundation for understanding the health of
 239 these sites, but as a primary prey source for the imperiled fishes PIBO was designed to monitor,
 240 an understanding of the macroinvertebrates of the sites is key to achieving management goals.
 241 Initial efforts to evaluate the usefulness of macroinvertebrate O/E ratios as the biotic component of
 242 conceptual models of degradation failed to find biologically significant results (Irvine et al., 2015),
 243 highlighting the need for a more effective metric for measuring the health of macroinvertebrate
 244 communities at PIBO sites.

245 Here we use phylogenetic diversity metrics to quantify the health of macroinvertebrate at sites within
 246 the PIBO dataset, demonstrating the continued use of PIBO data to generate actionable insights for
 247 land managers. Phylogenetic diversity metrics that incorporate a statistical expectation of diversity
 248 (SES_{MPD} and SES_{MNTD}) are more sensitive than those that do not (Faith's PD), complementing
 249 the current PIBO focus on taxonomic metrics incorporating expectations (O/E; an expectation of
 250 taxa that may be present at the site derived through modeling). Below, we discuss the management
 251 implications of the drivers we have identified in the data.

4.1 Contrasting patterns of phylogenetic and taxonomic diversity

Ratios of Observed to Expected taxa at a site (O/E ratios) are modified measures of species richness that can be compared across sites and studies (Hawkins, 2006). These ratios depend on models of the Probability of taxon Capture (PC) to predict the Expected pool (E), which impacts their performance (Clarke, Wright, & Furse, 2003). O/E ratio performance varies depending on the predictor variables used to generate the expectation and the PC thresholds used to calculate E (Hawkins, 2006; Hawkins et al., 2000). The failure of previous studies to find associations between anthropogenic drivers and the structure of macroinvertebrate assemblages (at PIBO; Irvine et al., 2015), demonstrates how these dependencies limit the application of O/E ratios to locations where PC models have been developed and adequate predictive data can be collected. There are, however, strong conceptual links between the phylogenetic diversity metrics we use here and previous O/E approaches. Faith's PD augments observed species richness with information about the phylogenetic diversity of those species, and so information about their traits and evolutionary history (Faith, 1992). This added information about the kinds of species within a site is useful, and SES_{MPD} and SES_{MNTD} build upon this by incorporating an evolutionary expectation of what a randomly assembled ecological community might resemble (Kembel, 2009; Webb et al., 2002). By combining expectations of community structure with information about the kinds of species in an assemblage, we can pick apart the drivers of community assembly and change. As sites are degraded or restored, it is possible for the number of species within a site to remain constant while the species themselves change; phylogenetic diversity metrics are sensitive to such changes. Critically, in our phylogenetic approach the expected pool comes from the observed data alone: there is no need for comparison sites in order to generate a statistical expectation allowing application of this method in locations that lack PC models. This is a clear practical advantage in favor of the use of phylogenetic diversity metrics.

Phylogenetic diversity is often viewed as a proxy for functional diversity, under the assumption that species that are more distantly related to one another have fewer functional traits in common (Devictor et al., 2010; Mazel et al., 2014; Tucker, Davies, Cadotte, & Pearse, 2018). This is certainly true, at a coarse scale, for macroinvertebrates: species within the same genus often

resemble one-another, but many different families and orders are instantly recognizable. Because environmental stressors alter the structure of a community they also predictably alter its phylogenetic diversity (Burns & Strauss, 2011; Cadotte et al., 2008; Cadotte et al., 2012; Flynn et al., 2011; Letcher, 2009). By understanding the phylogenetic structure of healthy communities we can identify damaged communities. Previous work on *Daphnia* indicated that disturbed communities tend to be phylogenetically clustered (Helmus et al., 2010), which is thought to reflect the new environmental filtering pressure of the disturbance. However, disturbed PIBO sites (*i.e.*, those with low condition indices) are instead overdispersed, with high SES_{MNTD} scores. Undisturbed communities with clustered structures, like PIBO sites, occur in other invertebrate systems such as ants, where invasion results in a reduction in clustering as groups of specialists are lost (Lessard et al., 2009). Here, phylogenetic diversity of macroinvertebrate assemblages is measured within each site (locally) and tends to be clustered in less disturbed sites. Since healthy macroinvertebrate communities are clustered at PIBO sites, we suggest that SES_{MPD} and SES_{MNTD} are sensitive enough to pick up on the losses of specialists within these communities. This phylogenetic clustering most likely reflects the high level of habitat heterogeneity found at PIBO sites, with highly specific habitats that host specialists. In contrast, communities at disturbed sites tended to be more over-dispersed – they contain more distantly-related species. This pattern would be consistent with a larger number of invasive (novel, distantly related) or the loss of specialist species, although we do not explicitly test those hypotheses here. Landscape scale stressors that reduce site heterogeneity would result in a loss of close relatives, as they are more likely to be functionally similar (Cavender-Bares et al., 2009). The phylogenetic clustering observed likely further reflects the environmental heterogeneity of PIBO sites as assemblages with larger source pools are expected to be more clustered (Cavender-Bares, Keen, & Miles, 2006; Pearse, Jones, & Purvis, 2013; Swenson, Enquist, Pither, Thompson, & Zimmerman, 2006).

4.2 Management implications of phylogenetic diversity

Improving and maintaining stream health for macroinvertebrates, and thus the imperiled fishes that prey upon them, depends on identifying the actionable factors that most dramatically disrupt

these communities. The relationships between the different physical measures of habitat quality that PIBO collects and the metrics we calculated allow us to highlight factors that are most influential on the macroinvertebrate communities, and some that were surprisingly not. PIBO's site designation as 'managed' or 'reference' is not a significant driver of disruption for O/E, SES_{MPD} , or SES_{MNTD} (Fig. 3, coefficient values overlap 0). During the original study design and implementation phase of the PIBO program, reference sites were selected to encompass both the greatest spatial extent possible as well as the range of variation in habitat conditions resulting from natural disturbances (*e.g.*, wildlife, landslides). Furthermore, not all anthropogenic disturbances (*e.g.*, recreation) were considered in the initial selection of reference sites. As a result, the binary classification of managed/reference alone is apparently insufficient to distinguish between intact and impaired macroinvertebrate assemblages. In contrast, quantitative measures of sites' abiotic condition (in the form of actionable explanatory variables) provide clear evidence that management affects these drivers of macroinvertebrate assemblage. PIBO's overall abiotic site condition index, significantly correlates with O/E, SES_{MPD} , and SES_{MNTD} , but not Faith's PD. This metric is the only of the four diversity metrics that does not incorporate some expectation of diversity. This result demonstrates that a statistical expectation of diversity is a useful component for metrics and is consistent with results from previous studies of the PIBO dataset. Other abiotic factors such as roads in a reach and/or its surroundings are some of the strongest factors influencing community structure. Road construction increases fine sediment and adversely impacts macroinvertebrates (Wood & Armitage, 1997). Measures of fine sediment, specifically percent pool tail fines present, also reflect this relationship, driving overdispersion nearly as strongly, or in the case of SES_{MNTD} more so than, roads. Because SES_{MPD} reflects broader patterns across the entire phylogeny, this is consistent with habitat filtering of entire lineages on the basis of the percent of pool tail fines <6mm present, and the percent of pool tail fines <2mm present driving competition among close relatives, thus having a greater impact on SES_{MNTD} . The bankfull width of the reach and frequency of pools are about half as influential as the previously mentioned measures, indicating that overall stream structure has some impact on the communities present. The amount of Total Dissolved Solids (TDS) and percent vegetatively stable banks have a comparatively minor influence on community structure, supporting previous findings that these measures may have little impact on macroinver-

tebrates and should not be the focus of management actions (Mazeika, Sullivan, Watzin, & Hession, 2004; Timpano, Schoenholtz, Zipper, & Soucek, 2010). That TDS were not important in our model does not mean TDS do not influence macroinvertebrates (Clements & Kotalik, 2016), particularly as measurement of this variable may be limited by measuring at base flows (Cey, Rudolph, Parkin, & Aravena, 1998). Rather, measures of pool tail fines and roads appear to be better indicators of macroinvertebrate assemblage response to management activities within the PIBO dataset.

Surprisingly, the frequency of large wood had little influence on macroinvertebrate communities when other explanatory variables such as sedimentation are included. Large wood has been hypothesized to be an important moderator of human stressors, specifically sedimentation (Irvine et al., 2015), but does not appear to strongly impact assemblage structure in these data. This indicates that indirect effects of large wood have mostly been accounted for by other explanatory variables in our model. In general, these results suggest that stream macroinvertebrates are consistently sensitive to road density, overall stream structure, and sedimentation. To remediate degraded macroinvertebrate assemblages, managers should focus efforts on actionable factors, such as roads, fine sediment, and the number of pools per kilometer, that have a large influence on community structure.

Simply identifying overdispersed communities does not adequately prioritize sites for management purposes. Community response to disturbance must be driven by physical measures that are ‘actionable’ and thus may be affected by management practices. Our statistical models identified a number of drivers of diversity in the PIBO data that are not likely to be influenced by management actions but are still important determinants of community structure (see Fig. 3). Of these non-actionable variables, stream gradient, elevation, and latitude are important predictors of SES_{MNTD} and SES_{MPD} , and geographic ‘hotspots’ and ‘coldspots’ are visible for all variables in Figure 4. These variables have nearly double the influence on our models as both fine sediment and road density, reflecting latitudinal gradient in biodiversity (Hillebrand, 2004), aquatic insect sensitivity to temperature (Johnson & Jones, 2000), and known patterns of decreased diversity in aquatic communities at higher elevations (Altermatt, Seymour, & Martinez, 2013). Even though these variables strongly influence the model, they are a result of site location and thus cannot be addressed by land

managers. We suggest that, all other things being equal, those sites that are strongly impacted by these factors (*i.e.*, are at high latitudes and are hot), should be a lower management priority. Some managers may be able to address factors that we designate as non-actionable. For example, grazing is known to impact instream temperatures (Kovach, Muhlfeld, Al-Chokhachy, Ojala, & Archer, 2019), and thus altering grazing intensity could be used to alter temperature and thus help remediate the site. By accounting for what cannot be changed and acting on what can, managers will effectively and efficiently be able to apply these metrics.

5 Conclusion

O/E ratios has provided strong management insight for decades, but like all approaches it has strengths and disadvantages. Here we have shown how phylogenetic diversity metrics, which have less ecological data requirements, can provide meaningful insights into the kinds of species present in communities. In this system, we have found phylogenetic diversity, when combined with an analysis targeted at quantifying the potential benefits of management interventions, to be of great insight. We encourage others to experiment with these new metrics and approaches, in the hope that they will be of use in other monitoring programs of macroinvertebrates and other taxa.

Author's Contributions

WDP, WCS, and KMW conceived the ideas and designed methodology; WCS and AVW collected, managed, and provided insight about the data; KMW and WDP analysed the data; KMW and WDP led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Acknowledgments

We are grateful to anonymous reviewers, and the editorial board, for their help improving this manuscript. WDP, KMW, and the Pearse lab are funded by NSF ABI-1759965, NSF EF-1802605, and USDA Forest Service agreement 18-CS-11046000-041.

Data accessibility

Data are publicly available by request through the USDA Forest Service PACFISH/INFISH Biological Opinion Monitoring Program (<https://www.fs.usda.gov/detail/r4/landmanagement/resourcemanagement/>) and the code to conduct analyses is available in the supplemental information.

Bibliography

- Altermatt, F., Seymour, M., & Martinez, N. (2013). River network properties shape α -diversity and community similarity patterns of aquatic insect communities across major drainage basins. *Journal of biogeography*, 40(12), 2249–2260.
- Balloch, D., Davies, C., & Jones, F. (1976). Biological assessment of water quality in three british rivers: The north esk (scotland), the ivel (england) and the taf (wales). *Water Pollution Control (UK)*.
- Bartoń, K. (2018). *Mumin: Multi-model inference*. R package version 1.42.1.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bivand, R., Keitt, T., & Rowlingson, B. (2018). *Rgdal: Bindings for the 'geospatial' data abstraction library*. R package version 1.3-6.
- Bivand, R., & Rundel, C. (2018). *Rgeos: Interface to geometry engine - open source ('geos')*. R package version 0.4-2.
- Bonada, N., Prat, N., Resh, V. H., & Statzner, B. (2006). Developments in aquatic insect biomonitoring: A comparative analysis of recent approaches. *Annual review of entomology*, 51, 495–523.
- Bossio, D., Geheb, K., & Critchley, W. (2010). Managing water by managing land: Addressing land degradation to improve water productivity and rural livelihoods. *Agricultural water management*, 97(4), 536–542.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference. *Sociological methods & research*, 33(2), 261–304.

- 414 Burns, J. H., & Strauss, S. Y. (2011). More closely related species are more ecologically similar in
415 an experimental test. *Proceedings of the National Academy of Sciences of the United States*
416 *of America*, 108(13), 5302–5307.
- 417 Butchart, S. H. M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J. P. W., Almond, R. E. A.,
418 ... Watson, R. (2010). Global biodiversity: Indicators of recent declines. *Science*, 328(5982),
419 1164–1168.
- 420 Cadotte, M. W., Cardinale, B. J., & Oakley, T. H. (2008). Evolutionary history and the effect of
421 biodiversity on plant productivity. *Proceedings of the National Academy of Sciences*, 105(44),
422 17012–17017.
- 423 Cadotte, M. W., Dinnage, R., & Tilman, D. (2012). Phylogenetic diversity promotes ecosystem
424 stability. *Ecology*, 93(8s), S223–S233.
- 425 Cadotte, M. W., Hamilton, M. A., & Murray, B. R. (2009). Phylogenetic relatedness and plant
426 invader success across two spatial scales. *Diversity and Distributions*, 15(3), 481–488.
- 427 Cadotte, M. W., & Tucker, C. M. (2018). Difficult decisions: Strategies for conservation prioritization
428 when taxonomic, phylogenetic and functional diversity are not spatially congruent. *Biological*
429 *conservation*, 225, 128–133.
- 430 Cairns, J., & Pratt, J. R. (1993). A history of biological monitoring using benthic macroinverte-
431 brates. *Freshwater biomonitoring and benthic macroinvertebrates*, 10, 27.
- 432 Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., ... Naeem, S.
433 (2012). Biodiversity loss and its impact on humanity. *Nature*, 486(7401), 59–67.
- 434 Cavender-Bares, J., Keen, A., & Miles, B. (2006). Phylogenetic structure of floridian plant commu-
435 nities depends on taxonomic and spatial scale. *Ecology*, 87(7), S109–S122.
- 436 Cavender-Bares, J., Kozak, K., Fine, P. V. A., & Kembel, S. W. (2009). The merging of community
437 ecology and phylogenetic biology. *Ecology letters*, 12, 693–715.
- 438 Cey, E. E., Rudolph, D. L., Parkin, G. W., & Aravena, R. (1998). Quantifying groundwater discharge
439 to a small perennial stream in southern ontario, canada. *Journal of Hydrology*, 210(1), 21–37.
- 440 Al-Chokhachy, R., Black, T. A., Thomas, C., Luce, C. H., Rieman, B., Cissel, R., ... Kershner, J. L.
441 (2016). Linkages between unpaved forest roads and streambed sediment: Why context matters

in directing road restoration: Unpaved roads and streambed sediment linkages. *Restoration Ecology*, 24(5), 589–598.

Al-Chokhachy, R., Ray, A. M., Roper, B. B., & Archer, E. (2013). Exotic plant colonization and occupancy within riparian areas of the interior columbia river and upper missouri river basins, USA. *Wetlands*, 33(3), 409–420.

Al-Chokhachy, R., Roper, B. B., & Archer, E. K. (2010). Evaluating the status and trends of physical stream habitat in headwater streams within the interior columbia river and upper missouri river basins using an index approach. *Transactions of the American Fisheries Society. Symposium* 48, 139(4), 1041–1059.

Al-Chokhachy, R., Roper, B. B., Archer, E. K., & Miller, S. (2011). Quantifying the extent of and factors associated with the temporal variability of physical stream habitat in headwater streams in the interior columbia river basin. *Transactions of the American Fisheries Society*, 140(2), 399–414.

Clarke, R. T., Wright, J. F., & Furse, M. T. (2003). RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological modelling*, 160(3), 219–233.

Clements, W. H., & Kotalik, C. (2016). Effects of major ions on natural benthic communities: An experimental assessment of the us environmental protection agency aquatic life benchmark for conductivity. *Freshwater Science*, 35(1), 126–138.

Coles-Ritchie, M. C., Roberts, D. W., Kershner, J. L., & Henderson, R. C. (2007). Use of a wetland index to evaluate changes in riparian vegetation after livestock exclusion. *Journal of the American Water Resources Association*, 43(3), 731–743.

Craven, D., Eisenhauer, N., Pearse, W. D., Hautier, Y., Isbell, F., Roscher, C., ... Manning, P. (2018). Multiple facets of biodiversity drive the diversity-stability relationship. *Nature ecology & evolution*, 2(10), 1579–1587.

de Zwart, D., Dyer, S. D., Posthuma, L., & Hawkins, C. P. (2006). Predictive models attribute effects on fish assemblages to toxicity and habitat alteration. *Ecological Applications*, 16(4), 1295–1310.

Devictor, V., Mouillot, D., Meynard, C., Jiguet, F., Thuiller, W., & Mouquet, N. (2010). Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: The need for integrative conservation strategies in a changing world. *Ecology letters*, 13(8), 1030–1040.

Díaz, S., Purvis, A., Cornelissen, J. H. C., Mace, G. M., Donoghue, M. J., Ewers, R. M., ... Pearse, W. D. (2013). Functional traits, the phylogeny of function, and ecosystem service vulnerability. *Ecology and evolution*, 3(9), 2958–2975.

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological conservation*, 61(1), 1–10.

Feld, C. K., Birk, S., Bradley, D. C., Hering, D., Kail, J., Marzin, A., ... Friberg, N. (2011). Chapter three - from natural to degraded rivers and back again: A test of restoration ecology theory and practice. In G. Woodward (Ed.), *Advances in ecological research* (Vol. 44, pp. 119–209). Academic Press.

Flynn, D. F. B., Mirotchnick, N., Jain, M., Palmer, M. I., & Naeem, S. (2011). Functional and phylogenetic diversity as predictors of biodiversity–ecosystem-function relationships. *Ecology*, 92(8), 1573–1581.

Friberg, N., Bonada, N., Bradley, D. C., Dunbar, M. J., Edwards, F. K., Grey, J., ... Woodward, G. (2011). Biomonitoring of human impacts in freshwater ecosystems: The good, the bad and the ugly. In G. Woodward (Ed.), *Advances in ecological research* (Vol. 44, pp. 1–68). Academic Press.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis*. Taylor & Francis.

Grueber, C. E., Nakagawa, S., Laws, R. J., & Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: Challenges and solutions. *Journal of evolutionary biology*, 24(4), 699–711.

Hadfield, J. D., Wilson, A. J., Garant, D., Sheldon, B. C., & Kruuk, L. E. B. (2010). The misuse of BLUP in ecology and evolution. *The American naturalist*, 175(1), 116–125.

Hämäläinen, H., Aroviita, J., Jyväsjärvi, J., & Kärkkäinen, S. (2018). Dangerous relationships: Biases in freshwater bioassessment based on observed to expected ratios. *Ecological Applications*, 28(5), 1260–1272.

- Hargett, E. G., ZumBerge, J. R., Hawkins, C. P., & Olson, J. R. (2007). Development of a RIVPACS-type predictive model for bioassessment of wadeable streams in Wyoming. *Ecological indicators*, 7(4), 807–826.
- Hastie, T., & Efron, B. (2013). *Lars: Least angle regression, lasso and forward stagewise*.
- Hawkins, C. P. (2006). Quantifying biological integrity by taxonomic completeness: Its utility in regional and global assessments. *Ecological applications: a publication of the Ecological Society of America*, 16(4), 1277–1294.
- Hawkins, C. P., Norris, R. H., Hogue, J. N., & Feminella, J. W. (2000). Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological applications: a publication of the Ecological Society of America*, 10(5), 1456–1477.
- Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals clock-like speciation and diversification. *Molecular biology and evolution*, msv037.
- Hellawell, J. M. (2012). *Biological indicators of freshwater pollution and environmental management*. Springer Science & Business Media.
- Helmus, M. R., Keller, W. B., Paterson, M. J., Yan, N. D., Cannon, C. H., & Rusak, J. A. (2010). Communities contain closely related species during ecosystem disturbance. *Ecology letters*, 13(2), 162–174.
- Henderson, R. C., Archer, E. K., Bouwes, B. A., Coles-Ritchie, M. S., & Kershner, J. L. (2005). *PIBO seven year status report 1998-2004*. US Forest Service.
- Hijmans, R. J. (2018). *Raster: Geographic data analysis and modeling*. R package version 2.8-4.
- Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. *The American naturalist*, 163(2), 192–211.
- Hough-Snee, N., Roper, B. B., Wheaton, J. M., & Lokteff, R. L. (2015). Riparian vegetation communities of the American Pacific Northwest are tied to multi-scale environmental filters. *River research and applications*, 31(9), 1151–1165.
- IPBES. (2018). The IPBES assessment report on land degradation and restoration. Bonn, Germany: Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

- Irvine, K. M., Miller, S. W., Al-Chokhachy, R. K., Archer, E. K., Roper, B. B., & Kershner, J. L. (2015). Empirical evaluation of the conceptual model underpinning a regional aquatic long-term monitoring program using causal modelling. *Ecological indicators*, 50, 8–23.
- Isaac, N. J. B., & Pearse, W. D. (2018). The use of EDGE (evolutionary distinct globally endangered) and EDGE-Like metrics to evaluate taxa for conservation. In R. A. Scherson & D. P. Faith (Eds.), *Phylogenetic diversity: Applications and challenges in biodiversity science* (pp. 27–39). Cham: Springer International Publishing.
- Isaac, N. J., Turvey, S. T., Collen, B., Waterman, C., & Baillie, J. E. (2007). Mammals on the edge: Conservation priorities based on threat and phylogeny. *PloS one*, 2(3), e296.
- Izsáki, J., & Papp, L. (1995). Application of the quadratic entropy indices for diversity studies of drosophilid assemblages. *Environmental and ecological statistics*, 2(3), 213–224.
- Johnson, S. L., & Jones, J. A. (2000). Stream temperature responses to forest harvest and debris flows in western cascades, oregon. *Canadian journal of fisheries and aquatic sciences. Journal canadien des sciences halieutiques et aquatiques*, 57(S2), 30–39.
- Kahle, D., & Wickham, H. (2013). Ggmap: Spatial visualization with ggplot2. *The R Journal*, 5(1).
- Kembel, S. W. (2009). Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. *Ecology letters*, 12(9), 949–960.
- Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., ... Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26(11), 1463–1464.
- Kershner, J. L., Archer, E. K., Coles-Ritchie, M., Cowley, E. R., Henderson, R. C., Kratz, K., ... Vinson, M. R. (2004). *Guide to effective monitoring of aquatic and riparian resources* (tech. rep. No. RMRS-GTR-121). U.S. Department of Agriculture, Rocky Mountain Research Station.
- Kovach, R. P., Muhlfeld, C. C., Al-Chokhachy, R., Ojala, J. V., & Archer, E. K. (2019). Effects of land use on summer thermal regimes in critical salmonid habitats of the pacific northwest. *Canadian Journal of Fisheries and Aquatic Sciences*, 76(5), 753–761.
- Lessard, J.-P., Fordyce, J. A., Gotelli, N. J., & Sanders, N. J. (2009). Invasive ants alter the phylogenetic structure of ant communities. *Ecology*, 90(10), 2664–2669.

- 555 Letcher, S. G. (2009). Phylogenetic structure of angiosperm communities during tropical forest
556 succession. *Proceedings of the Royal Society B: Biological Sciences*, 27, 97–104.
- 557 Macadam, C. R., & Stockan, J. A. (2015). More than just fish food: Ecosystem services provided by
558 freshwater insects: Ecosystem services and freshwater insects. *Ecological entomology*. CEEBA
559 Publications Série II, Vol. 69, 40, 113–123.
- 560 Mayfield, M. M., & Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylo-
561 genetic structure of communities. *Ecology letters*, 13(9), 1085–1093.
- 562 Mazeika, S., Sullivan, P., Watzin, M. C., & Hession, W. C. (2004). Understanding stream geomorphic
563 state in relation to ecological integrity: Evidence using habitat assessments and macroinver-
564 tebrates. *Environmental management*, 34(5), 669–683.
- 565 Mazel, F., Guilhaumon, F., Mouquet, N., Devictor, V., Gravel, D., Renaud, J., ... Mouillot, D.,
566 et al. (2014). Multifaceted diversity–area relationships reveal global hotspots of mammalian
567 species, trait and lineage diversity. *Global ecology and biogeography*, 23(8), 836–847.
- 568 Meals, D. W., Dressing, S. A., & Davenport, T. E. (2010). Lag time in water quality response to
569 best management practices: A review. *Journal of environmental quality*, 39(1), 85–96.
- 570 Melland, A. R., Fenton, O., & Jordan, P. (2018). Effects of agricultural land management changes
571 on surface water quality: A review of meso-scale catchment research. *Environmental science*
572 *& policy*, 84, 19–25.
- 573 Menuz, D. R., & Kettenring, K. M. (2013). The importance of roads, nutrients, and climate for
574 invasive plant establishment in riparian areas in the northwestern united states. *Biological*
575 *invasions*, 15(7), 1601–1612.
- 576 Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: Synthesis*. Island
577 Press, Washington, DC.
- 578 Moya, N., Hughes, R. M., Domínguez, E., Gibon, F.-M., Goitia, E., & Oberdorff, T. (2011).
579 Macroinvertebrate-based multimetric predictive models for evaluating the human impact on
580 biotic condition of bolivian streams. *Ecological indicators*, 11(3), 840–847.
- 581 Newbold, T., Hudson, L. N., Hill, S. L. L., Contu, S., Lysenko, I., Senior, R. A., ... Collen, B.,
582 et al. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, 520(7545),
583 45–50.

- Pearse, W. D., Chase, M. W., Crawley, M. J., Dolphin, K., Fay, M. F., Joseph, J. A., ... Purvis, A. (2015). Beyond the edge with edam: Prioritising british plant species according to evolutionary distinctiveness, and accuracy and magnitude of decline. *PLOS ONE*, 10(5), 1–18.
- Pearse, W. D., Cadotte, M. W., Cavender-Bares, J., Ives, A. R., Tucker, C. M., Walker, S. C., & Helmus, M. R. (2015). Pez: Phylogenetics for the environmental sciences. *Bioinformatics*, 31(17), 2888–2890.
- Pearse, W. D., Purvis, A., Cavender-Bares, J., & Helmus, M. R. (2014). Metrics and models of community phylogenetics. In *Modern phylogenetic comparative methods and their application in evolutionary biology* (pp. 451–464). Springer.
- Pearse, W. D., Jones, A., & Purvis, A. (2013). Barro colorado island’s phylogenetic assemblage structure across fine spatial scales and among clades of different ages. *Ecology*, 94(12), 2861–2872.
- Prather, C. M., Pelini, S. L., Laws, A., Rivest, E., Woltz, M., Bloch, C. P., ... Joern, A. (2013). Invertebrates, ecosystem services and climate change. *Biological reviews of the Cambridge Philosophical Society*, 88(2), 327–348.
- Procheş, Ş., Wilson, J. R. U., Richardson, D. M., & Rejmánek, M. (2008). Searching for phylogenetic pattern in biological invasions. *Global Ecology and Biogeography*, 17(1), 5–10.
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Resh, V. H., & Rosenberg, D. M. (1993). *Freshwater biomonitoring and benthic macroinvertebrates*.
- Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 6(1), 15–32.
- Roper, B. B., Jarvis, B., & Kershner, J. L. (2007). The role of natural vegetative disturbance in determining stream reach characteristics in central idaho and western montana. *Northwest science: official publication of the Northwest Scientific Association*, 81(3), 224–238.
- Schmera, D., Heino, J., Podani, J., Erős, T., & Dolédec, S. (2017). Functional diversity: A review of methodology and current knowledge in freshwater macroinvertebrate research. *Hydrobiologia*, 787(1), 27–44.

- Smith, M. J., Kay, W. R., Edward, D. H. D., Papas, P. J., et al. (1999). AusRivAS: Using macroinvertebrates to assess ecological condition of rivers in western australia. *Freshwater*.
- Strauss, S. Y., Webb, C. O., & Salamin, N. (2006). Exotic taxa less related to native species are more invasive. *Proceedings of the National Academy of Sciences*, 103(15), 5841–5845.
- Swenson, N. G., Enquist, B. J., Pither, J., Thompson, J., & Zimmerman, J. K. (2006). The problem and promise of scale dependency in community phylogenetics. *Ecology*, 87(10), 2418–2424.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 58(1), 267–288.
- Timpano, A. J., Schoenholtz, S. H., Zipper, C. E., & Soucek, D. J. (2010). Isolating effects of total dissolved solids on aquatic life in central appalachian coalfield streams. In *Proc. 2010 national meeting of the american society of mining and reclamation, pittsburgh, PA* (pp. 1284–1302).
- Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., ... Mazer, F. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological reviews of the Cambridge Philosophical Society*, 92(2), 698–715.
- Tucker, C. M., Davies, T. J., Cadotte, M. W., & Pearse, W. D. (2018). On the relationship between phylogenetic diversity and trait diversity. *Ecology*.
- Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., ... Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*, 467(7315), 555–561.
- Webb, C. O., Ackerly, D. D., McPeck, M. A., & Donoghue, M. J. (2002). Phylogenies and community ecology. *Annual review of ecology and systematics*, 33(1), 475–505.
- Wood, P. J., & Armitage, P. D. (1997). Biological effects of fine sediment in the lotic environment. *Environmental management*, 21(2), 203–217.
- Wright, J. F., Sutcliffe, D. W., & Furse, M. T. (2000). An introduction to RIVPACS. In J. F. Wright, D. W. Sutcliffe, & M. T. Furse (Eds.), *Assessing the biological quality of freshwater: RIVPACS and other techniques* (pp. 1–24). FBA Special Publications. Ambleside, UK: Freshwater Biological Association.
- Wright, J. F. (1995). Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. *Austral ecology*, 20(1), 181–197.

641 Yates, A. G., Brua, R. B., Culp, J. M., Chambers, P. A., & Wassenaar, L. I. (2014). Sensitiv-
 642 ity of structural and functional indicators depends on type and resolution of anthropogenic
 643 activities. *Ecological indicators*, 45, 274–284.

644 Figures

Figure 1: The distribution of the 1667 PIBO sampling sites across the Interior Columbia and Upper Missouri river basins. Sites include 1409 standard sites that are sampled every five years, 100 pilot sites, and 158 contracted or special program sites.

Figure 2: The relationships between condition index and SES_{MNTD} , SES_{MPD} , Faith's PD, and O/E, showing both managed and reference sites. The results of a linear regression for each of these metrics indicated that site condition index score explained 12.7%, 9.8%, 0.03%, and 18.6% of the variance respectively. Site condition index score was significantly correlated with O/E ($F_{1,4580} = 1049$, $p < 0.0001$, $r^2 = 0.186$), SES_{MPD} ($F_{1,4588} = 504.5$, $p < 0.0001$, $r^2 = 0.099$), and SES_{MNTD} ($F_{1,4588} = 641.2$, $p < 0.0001$, $r^2 = 0.123$), but not Faith's PD ($F_{1,4631} = 1.51$, $p = 0.220$, $r^2 = 0.0003$). Thus O/E is a better metric of the condition index than our phylogenetic diversity indices, which is to be expected as O/E uses additional data to calculate its expectation, but that its performance is comparable to that of SES_{MNTD} and SES_{MPD} . However, designation of a site as reference significantly explained only 2.9%, 1.4%, 0.6%, and 5.2% of variance for all metrics respectively (SES_{MNTD} ($F_{1,4978} = 141.3$, $p < 0.0001$, $r^2 = 0.028$), SES_{MPD} ($F_{1,4978} = 67.44$, $p < 0.0001$, $r^2 = 0.013$), Faith's PD ($F_{1,5031} = 33.91$, $p < 0.0001$, $r^2 = 0.007$), and O/E ($F_{1,4975} = 272.2$, $p < 0.0001$, $r^2 = 0.052$)). This indicates that reference site conditions are not necessarily good benchmarks for macroinvertebrate communities.

Figure 3: Plots showing the relative importance of each explanatory variable considered in the model for each metric considered. Coefficient values further from zero (either in the positive or negative direction) indicate that the variable has a greater impact on the overall model. For example, the percent of pool tail fines < 2mm present has nearly twice the influence on the SES_{MNTD} model as pool frequency. Explanatory variables are indicated on the left and are colored by whether or not they are considered potentially actionable by a manager attempting to address habitat degradation. This indicates there are consistently influential actionable explanatory variables.

Figure 4: Plots showing site-level variation across each metric, highlighting sites with unusually high or low condition index values given the variation within the dataset for each explanatory variable. In all cases, sites in red should be management priorities. In A (SES_{MNTD}) and B (SES_{MPD}), sites in blue have more positive values, indicating more distantly-related species are found at a site, and values in red have more closely-related species at a site (more negative values), than would be expected given site conditions (management, disturbance, latitude, and all other variables in our models). Again, given clustering (more negative) is found at more disturbed sites, sites in red should be management priorities. In C (Faith's PD), sites in red have less phylogenetic diversity, and sites in blue more diversity, than would be expected by chance, and thus sites in red are management priorities. Finally, in D (O/E) sites in red have fewer species observed than expected, while sites in blue have more species than the expectation. We emphasize that, as discussed in the text, these estimates should not be used to generalize management decisions to other sites or data, and should only be considered within the context of this modeling exercise.