Faculty Scholarship Collection

Accounting for non-native Brown Trout in biological assessments: Implications for selecting reference conditions

Mark A. Kirk[1,2,3] **and Scott A. Wissinger**[2,4]

[1]Department of Zoology and Physiology, University of Wyoming, 1000 East University Avenue, Laramie, Wyoming 82071 USA
[2]Biology and Environmental Science Departments, Allegheny College, 520 North Main Street, Meadville, Pennsylvania 16335 USA

Abstract: The efficacy of assessments that evaluate biological integrity can be improved by accounting for the ecological processes that influence assemblage composition. Many studies have emphasized that bioassessments need to account for natural environmental gradients, but there is little consensus on how bioassessments should account for the impacts of non-native species. In particular, non-native trout species have been introduced into many high-quality streams that probably meet the expected reference conditions for bioassessment in a given region. The goal of this study was to test whether the presence of large, piscivorous, non-native Brown Trout (*Salmo trutta*) at reference sites altered interpretations of taxonomic completeness indices (TCIs) based on fish assemblages. We used fish data from 215 sites in wadeable streams in northwestern Pennsylvania to compare the performance of 3 TCIs that used different modeling approaches to account for Brown Trout impacts. One model accounted for the presence of non-native Brown Trout as a covariate (covariate model), another model censored reference sites with non-native Brown Trout from the reference pool (censored model), and a final model used all reference sites without accounting for the presence of Brown Trout (unaccounted model). TCIs based on observed-to-expected ratios were able to distinguish reference conditions from altered conditions in the covariate and censored models. In contrast, the unaccounted index could not distinguish reference from altered conditions and was, thus, unable to accurately assess biological integrity, probably because large Brown Trout reduce native species richness. Our results provide a framework for how bioassessment practitioners can use different approaches to account for non-native species impacts, especially when considering which criteria are most important for defining reference conditions. Accounting for the effects of non-native species with these approaches should improve the ability of bioassessments designed to summarize the interactive effects of all potential human stressors on stream assemblages.
Key words: reference conditions, non-native species, taxonomic completeness, biological assessment, biological integrity

Aquatic ecologists use biological criteria to evaluate the health of aquatic ecosystems to address the aspirations of the federal US Clean Water Act of 1972 (Karr 1991, Adler 2012). Bioassessment tools, such as taxonomic completeness indices (TCIs) and multi-metric indices (MMIs), are designed to assess the multiple and interactive effects of land-use impacts and non-point stressors on aquatic ecosystems. These tools have been adapted for, and applied to, stream monitoring worldwide (Ruaro and Gubiani 2013). MMIs integrate taxonomic and functional attributes of assemblage structure into a numeric index (e.g., Angermeier and Karr 1994, Angermeier et al. 2000), whereas TCIs assess the intactness of an aquatic assemblage as measured by the observed-to-expected ratio (O/E) of taxa within an assemblage (Hawkins 2006, Meador and Carlisle 2009, Vander Laan and Hawkins 2014, Mazor et al. 2016). Bioassessment based on these indices typically relies on the establishment of reference conditions and the suite of taxa that we expect to see under these conditions (Hawkins et al. 2010). Thus, identifying appropriate reference conditions is critical to establishing effective bioassessments (Chessman and Royal 2004, Stoddard et al. 2006).

E-mail addresses: [3]kirkma18@gmail.com; [4]swissing@allegheny.edu

DOI: 10.1086/705918. Received 10 February 2019; Accepted 12 July 2019; Published online 18 September 2019.
Freshwater Science. 2019. 38(4):790–801.

Perhaps the most common challenge associated with the development of bioassessments is to separate human-induced impacts from natural environmental variation (Allan 2004, Southerland et al. 2007). TCIs attempt to account for natural environmental variation among reference sites to increase their ability to detect human-caused alterations. The majority of bioassessment studies identify reference sites based on some minimal level of landscape-level alteration (Hill et al. 2017), but other criteria at reach- or site-specific scales can be useful for identifying and excluding sites from the reference pool when multiple stressors are present (Chessman and Royal 2004, Ode et al. 2016). For example, methods that account for the potential effects of non-native or invasive species have rarely been considered during the development and testing of TCIs, despite acknowledgement that non-natives should be a reference-site criterion (Kennard et al. 2005, Meador and Carlisle 2009, Sánchez-Montoya et al. 2009, Mazor et al. 2016).

Freshwater ecosystems are among the most invaded on the planet, which has resulted in biodiversity loss and biotic homogenization (Rahel 2000, Leprieur et al. 2008, Strayer 2010). Ignoring potential interactions between native and non-native species could bias interpretations of biological integrity if non-native species alter local assemblages. Many of the most commonly introduced non-native fish species are large-bodied piscivores that can affect native fish species through predation or competition (Mitchell and Knouft 2009). These large-bodied non-native species have been introduced across the full spectrum of aquatic conditions, from centrarchid species (e.g., sunfish and bass) in warm-water streams to salmonid species in cold-water streams (Ross 1991, Rahel 2000, Scoppettone et al. 2005, Fausch 2008, Hermoso et al. 2011). Non-native species could bias assessments of assemblage completeness by inflating observed richness in some instances (Scott and Helfman 1999) or in other instances reducing the expected richness via antagonistic interactions with native species (Kirk et al. 2017).

The objective of this study was to assess the effects of large, piscivorous, and non-native Brown Trout (*Salmo trutta*) on the performance of fish TCIs. Brown Trout have been listed as 1 of the 30 most globally invasive fish species and negatively affect native fish assemblages around the world (reviewed in Budy and Gaeta 2017, see also McIntosh 2000, McDowall 2003, Zimmerman and Vondracek 2006, Sowersby et al. 2015, Turek et al. 2016). Brown Trout are relatively tolerant to different environmental conditions and landscape-level changes (McIntosh et al. 2011, Wagner et al. 2013) and therefore have the potential to interact with many different native assemblages (Simon and Townsend 2003). Importantly, Brown Trout appear to negatively modify native fish assemblages in heavily forested (i.e., reference) streams of our study region in northwestern Pennsylvania, which could confound bioassessments if Brown Trout streams are used as reference sites (see Kirk et al. 2017).

We used fish-based TCIs (O/E) to compare different approaches that account for the effects of non-native Brown Trout on native species in high-quality streams (i.e., reference conditions) and determine whether those approaches alter interpretations of biological integrity. We evaluated 3 approaches for modeling expected number of taxa (E) and estimating O/E values: 1) an approach that included references sites with Brown Trout in the reference-site pool and used them as a predictor variable to account for their effects on O/E values (covariate model), 2) an approach that had reference sites with Brown Trout present censored from the reference pool (censored model), and 3) an approach that grouped all reference sites without considering Brown Trout presence as a possible predictor variable (unaccounted model). We then compared the performance of the 3 indices based on bioassessment efficacy metrics (responsiveness, precision, sensitivity, and bias; Hawkins 2006, Vander Laan and Hawkins 2014, Mazor et al. 2016) and an assessment of the accuracy of species distribution models (area under the curve [AUC] values; Rose et al. 2016). We hypothesized that the unaccounted model would have lower assessment performance for all metrics compared with results from the censored model and covariate model. Specifically, we predicted O/E indices would be unable to distinguish reference conditions from altered conditions without accounting for the presence of large Brown Trout because they would reduce the occurrence of native species that would otherwise be found under reference conditions.

METHODS

Study region and data aggregation

Our fish assemblage dataset was composed of 215 unique sampling locations from wadeable streams in 4 basins in the Upper Allegheny River watershed, Pennsylvania, USA (French Creek, Oil Creek, Brokenstraw Creek, and direct tributaries of the Allegheny River; Fig. 1). Watershed area of our study streams ranged from 0.5 to 50 km^2. Small streams with watershed areas <10 km^2 are cold-water systems and, in our study region, represent the smallest streams that contain large-bodied, piscivorous Brown Trout (large-bodied and piscivorous defined as >150 mm; McIntosh 2000, Kirk et al. 2018). Medium streams are cool-water systems that range in watershed size from 10 to 50 km^2. The landscape is predominately a mix of low-intensity agriculture and secondary-growth forest, with elevation and forest cover increasing to the east (Kirk et al. 2017, 2018).

We sampled fish assemblages between 2005 and 2016 with a pulsed DC backpack electrofisher (Smith-Root Model 12; Smith-Root Company, Vancouver, Canada) from May through October during low-flow conditions. We conducted a fish survey at each site across a 150- to 250-m stream reach to collect data on both relative abundance and species presence/absence (see Kirk et al. 2017, 2018). First,

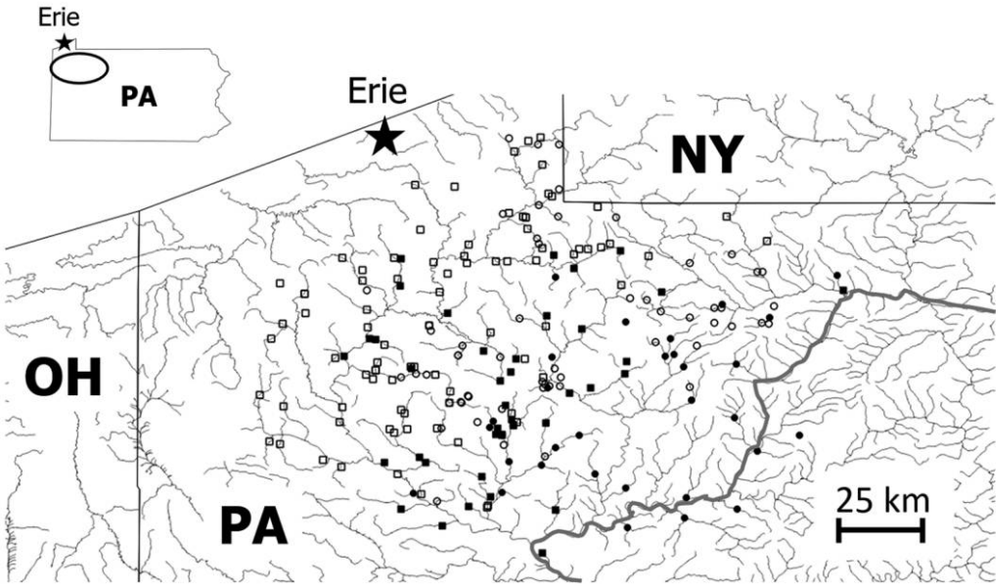

Figure 1. Map of the 215 sample sites within the Upper Allegheny River watershed of northwestern Pennsylvania. Closed circles = sites with Brown Trout present, open circles = sites with Brown Trout absent, closed squares = non-reference sites (medium/altered) with Brown Trout present, and open squares = non-reference sites with Brown Trout absent. The star indicates the location of Erie, Pennsylvania. The dark gray line denotes the mainstem of the Allegheny River.

we herded fish to an upstream barrier and collected a quantitative sample to enumerate individuals of all species (e.g., blocking seine net, riffle, etc.). This procedure was conducted either once (single-pass estimate in a 100-m reach) or twice (2-pass depletion estimate in a 50-m reach). We subsequently took a qualitative sample in an adjacent 100- to 150-m reach to ensure comprehensive sampling of all habitat types and document the presence of any additional species. All analyses used presence/absence data based on the combined quantitative and qualitative samples to provide an adequate sampling distance (150–250 m) for documenting species occurrence.

Environmental data consisted of landscape-based variables and water chemistry data. Landscape data were collected in Geographic Information Systems (ARCMAP by ERSI on ArcGIS Desktop 9.3.1; Redlands, California, USA) following methods described in Kirk et al. (2017). For each sample location we determined site latitude and longitude, upstream watershed land cover type (%), upstream riparian land cover within a 50-m buffer of the stream channel (%), stream elevation (m), upstream watershed area (km^2), downstream distance from each sample site to the mainstem of the respective sub-basin (km), and channel slope (m/km elevation drop from source to the sample location based on National Hydrography Dataset flow lines). We calculated land-cover types with either the 2006 (sites sampled during 2005–2011) or the 2012 (sites sampled during 2012–2016) National Land Cover Dataset. We organized the land-cover data into 2 categories: 1) natural land cover, which included all forest and wetland land-cover types; and 2) non-natural land cover, which included the combination of agricultural (e.g., row crops, pasture) and developed land use (e.g., urbanization). Alkalinity was the only water chemistry variable we used in our analyses.

O/E index construction and development

Construction of O/E indices requires identifying reference sites that are in excellent ecological condition. These reference sites allow the prediction of the probabilities of species occurrence at individual sites expected under no, or minimal, human-caused stress (Hawkins et al. 2000, Hawkins 2006). This set of reference sites was our training dataset. Sites were selected based on a set of reference score criteria. Minimum criteria for inclusion in the reference-site pool were that the summed upstream watershed natural cover and upstream riparian natural cover was ≥150% (200% maximum, Table 1, Fig. S1) and that a site's upstream watershed had minimal row crop agriculture (<15%) and urbanization cover (<5%). These specific criteria were either similar to or more conservative than least altered conditions used in 2 TCIs for eastern USA streams (Carlisle et al. 2008, Meador and Carlisle 2009). We also selected a set of altered sites based on scores that had combined watershed and riparian natural cover of <120%. We classified all remaining sites (scores of 120–150%) as medium-impact sites. We refer to the set of altered and medium impact sites as test sites.

Table 1. Means and standard deviations (SD) of environmental characteristics for 215 sites in the 4 land-use classes used in analyses. Note that reference site groups are defined based on whether they contain Brown Trout ($n = 88$; covariate and unaccounted models) or not ($n = 57$; censored model).

	Reference with LBT censored ($n = 57$)		Reference with LBT included ($n = 88$)		Medium impact ($n = 90$)		Altered sites ($n = 37$)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Reference score (%)	168.6	15.4	170.2	15.0	135.8	10.2	104.6	15.1
Natural cover (%)	78.3	9.8	79.6	10.2	60.4	8.7	44.6	8.8
Riparian cover (%)	90.2	7.2	90.5	6.6	75.7	7.0	61.1	8.1
Latitude (°)	41.8	0.1	41.7	0.1	41.7	0.2	41.8	0.2
Longitude (°)	−79.7	0.2	−79.7	0.2	−79.9	0.2	−80.0	0.2
Watershed area (km^2)	8.0	7.6	11.3	9.9	12.6	9.2	11.9	9.5
Elevation (m)	423.0	39.0	411.4	41.8	392.6	38.7	377.8	40.2
Channel slope (m/km)	24.0	12.4	21.2	11.0	15.5	8.1	14.0	8.8
Distance (km)	10.6	10.1	10.6	9.9	13.2	11.1	14.1	11.8
Alkalinity	49.3	27.4	49.3	26.2	59.9	23.8	76.8	31.6

The number of reference sites used in a model differed between the censored model and the other models. For the 2 models that did not censor sites with Brown Trout present from the reference-site pool (the covariate and unaccounted models), 88 of the 215 sites (41%) were considered to be in reference condition, whereas the censored model had only 57 (27%) reference sites (i.e., 31 reference sites with Brown Trout present were removed). This sample size difference resulted in reduced representation of some environmental conditions for the reference dataset of the censored model, which provided an opportunity to evaluate how different reference samples could confound model comparisons (e.g., Ode et al. 2016). The 2 models that included Brown Trout reference conditions incorporated sites with larger watershed areas, lower elevations, and lower channel slopes than the censored model (Table 1). This difference occurred because Brown Trout populations are less likely to occur in small, high-gradient streams in this study region (see Kirk et al. 2018). Reference sites for all models tended to be farther east in high-elevation streams with relatively low alkalinities and high channel slopes (Table 1; Kocovsky and Carline 2005).

We excluded species that were present ≥90 or <10% of reference sites from all analyses to avoid effects of ubiquitous and rare species (Hawkins et al. 2000, Vander Laan and Hawkins 2014, Mazor et al. 2016). This exclusion resulted in a total of 21 native fish species that we used in development of all 3 O/E indices (see Table S1 for full species list; Brown Trout were excluded from all analyses). The 21 native species represent all major taxonomic families and ecological trait guilds, and all are widely distributed across our study region (Stauffer et al. 2016, Kirk et al. 2017).

Methods for O/E index development have been well described elsewhere (Hawkins et al. 2000, Van Sickle et al. 2006), which we briefly summarize here. The 1st step was to identify clusters of reference sites based on their biological similarity. We used Sørensen's dissimilarities of presence/absence data on the 21 native species to create 1 dendrogram for the censored model ($n = 57$) with the unweighted pair–group arithmetic averaging algorithm, and then 1 dendrogram for the uncensored and covariate models via the same process ($n = 88$; *cluster* package; R version 3.2.3; R Foundation for Statistical Computing, Vienna, Austria). We used these dendrograms to identify reference-site clusters that maximized within-cluster site similarity while ensuring that clusters had ≥5 sites (see Figs S2–S3; Van Sickle et al. 2006, Carlisle et al. 2008). We then estimated the probability of capturing (Pc) a species at a test site, assuming it was in reference condition, as the probability of a site belonging to each cluster multiplied by the frequency of that species within a cluster summed across all clusters (Wright 1995).

The 2nd step was to use multiple discriminant analysis (MDA) to select environmental variables that best estimated group (cluster) membership. Models for estimating group membership were based on 8 potential variables: latitude, longitude, elevation, watershed area, distance to mainstem, channel slope, alkalinity, and the presence of large, piscivorous Brown Trout (LBT; binary variable). All variables, except for the presence of Brown Trout, are minimally influenced by human activities at our study sites. None of the variables were strongly correlated (all $|r| < 0.57$). Watershed area and distance to mainstem were log-transformed to meet the assumptions of normality and homogeneity of variances for the MDA.

To determine the best model for predicting group membership and subsequent Pcs, we used custom R scripts and the best subsets approach (Van Sickle et al. 2006). We

identified the best models as those that were most parsimonious (fewest predictors to avoid model overfitting), accurate (high correct classification rates), and precise (low root mean square errors). We repeated the best subsets evaluation independently for all 3 modeling approaches. One best subsets model was determined for the censored model with 57 reference sites. For the models that included Brown Trout reference sites ($n = 88$), 1 best subsets evaluation was made with all 8 candidate predictor variables that would directly account for Brown Trout presence if selected in the final model (covariate model). Another separate best subsets evaluation was then done with only the 7 physical–chemical variables (unaccounted model).

Previous development of O/E indices suggests that including species with a Pc above a certain threshold can affect index performance (Hawkins et al. 2000, Hawkins 2006, Vander Laan and Hawkins 2014, Mazor et al. 2016). Most studies recommend a Pc threshold of 0.5 for inclusion or exclusion of species in a model, but we applied a species-specific threshold for each model described above because of the high species omission rates observed for rare taxa in TCIs (see Rose et al. 2016). Species-specific Pc thresholds were determined by calculating Cohen's Kappa statistic based on the observed presence/absence data and then estimating Pc data across all possible Pc values between 0 and 1 in 0.05 increments. The Kappa statistic provides a measure of correct classification after accounting for potential chance agreement, and the Pc that maximized the Kappa statistic for each species was selected for each index (see Canning 2018). We then predicted E at a site as the sum of all species that had Pc values above their respective thresholds. The number of observed taxa (O) was the number of species observed at each site that were predicted to occur based on their respective Pc threshold. Sites with lower O/E ratios (e.g., <1) are typically inferred to be taxonomically incomplete compared with the assemblages expected under reference conditions.

Index performance

To evaluate performance of the 3 different O/E indices, we first regressed O (y-axis) against E (x-axis) for data from all sites. We compared the coefficient of determination (R^2), the slope, and the intercept for each modeling approach with expectations from a 1:1 line (i.e., a perfect model fit) with simple linear regression (*lm* function; R version 3.2.3; R Foundation for Statistical Computing, Vienna, Austria). R^2 estimates the congruency between O and E and the slope and intercept assess model bias and consistency (Piñeiro et al. 2008).

We evaluated the accuracy of the 21 species distributions predicted from each of the 3 modeling approaches based on AUC estimates. Capture probabilities from the models were used to compute AUC values, which ranged from 0.5 to 1.0. Values of 0.5 indicate very weak predictive performance and 1.0 indicates perfect performance. Values >0.7 have been interpreted to indicate good model performance (Rose et al. 2016). We compared AUC values both statistically and descriptively. First, we used analysis of variance (ANOVA) to test if the 3 modeling approaches differed in mean AUC values. Second, we counted the number of species with AUC values >0.7 (i.e., good performance). We were particularly interested in testing whether the censored model had different predictive accuracy because it had a smaller, different reference pool sample than the covariate and unaccounted models.

We also used 4 additional criteria (responsiveness, bias, sensitivity, and precision) for assessing index performance that have been used in previous bioassessment studies (Hawkins 2006, Hawkins et al. 2010, Vander Laan and Hawkins 2014, Mazor et al. 2016). Responsiveness is a measure of how much bioassessment scores change with site alteration, and we evaluated it with a 1-way ANOVA (R version 3.2.3; available from: https://cran.r-project.org/bin/windows/base/old/3.2.3/) that tested for differences in mean O/E index scores among the 4 land-use classes (altered, medium, and reference sites with and without Brown Trout). We then used Tukey's post-hoc honestly significant difference (HSD) tests to test for specific differences in index scores between altered and reference sites. We evaluated bias with a multiple regression model (*lm* function) to determine if any remaining variation in the index scores could be attributed to the predictor variables included in the models. The 3rd criterion, sensitivity, was defined as the proportion of altered sites that had O/E values below the 10th percentile of reference site values.

The 4th criterion, precision, measures the amount of variability in reference site values and was quantified as the standard deviation (SD) in reference site O/E values. Application of bioassessment models to data from reference sites should theoretically produce small deviations in assessment scores that would only be associated with sampling error. The overall performance of O/E indices is best assessed by applying the model to an independent set of reference sites, known as a validation sample (Hawkins et al. 2000, Hawkins 2006, Van Sickle et al. 2006). Small sample sizes in our study prevented us from setting aside a validation set. Thus, we validated the performance of each O/E index by applying the model to each respective set of reference sites.

Species occurrence frequencies

We compared differences in responses of individual species between the 3 modeling approaches by calculating the observed frequencies (O_F) and expected frequencies (E_F) for the 21 native species. This comparison examined which species occurred more or less frequently than expected based on reference conditions (Carlisle et al. 2008, Meador and Carlisle 2009). O_F was calculated as the sum of observed species occurrences across all non-reference

sites that were predicted to occur with Pc values greater than each species' respective threshold. E_F was calculated as the sum of predicted species occurrences with Pc values greater than each respective threshold across the same set of non-reference sites. We then calculated the $O_F : E_F$ ratio to evaluate each species' sensitivity to alterations at non-reference sites (i.e., sensitivity to deviations from reference conditions). We defined species with $O_F : E_F \geq 1.25$ as tolerant increasers (i.e., more frequently observed at stream sites with altered conditions), and species with $O_F : E_F \leq 0.75$ as sensitive decreasers (i.e., less frequently observed with increasing alteration). If all models predicted species distributions similarly, we expected that $O_F : E_F$ ratios would not differ among the 3 modeling approaches.

RESULTS

The reference clusters for all 3 models were similarly associated with gradients of stream size and elevation. For the covariate and unaccounted models that included Brown Trout reference sites ($n = 88$), reference clusters were strongly linked to stream size. Of 7 total clusters, 100% of sites in 4 clusters had watershed areas <10 km^2, and 91% of sites in the remaining 3 clusters had watershed areas of 10 to 50 km^2. The 4 small-stream clusters mainly differentiated low-elevation streams in the western portion of our study region from high-elevation, high-gradient streams in the eastern portion of our study region. The 3 large-stream clusters mainly differed in the presence of Brown Trout. In 1 large-stream cluster, 86% of sites (12 of 14) had no LBT, whereas in the 2 other large-stream clusters, 63% and 75% (10 of 16 and 12 of 16, respectively) of sites had LBT. Reference clusters in the censored model ($n = 57$) were strongly associated with stream size, elevation, and longitude, similar to the reference clusters of the 2 uncensored models (see Figs S2 and S3 for more details).

The best subsets process from the MDA resulted in a similar suite of predictor variables for the 3 modeling approaches. Reference clusters used in the covariate model were best predicted by an MDA model that contained alkalinity, watershed area, distance to mainstem, latitude, longitude, and the presence of LBT. This was the only model ranked in the top 10% of all possible covariate models for both accuracy and precision, and correct classification rates (i.e., accuracy) were 6% higher than the competing models with slightly higher precision. In the unaccounted model, reference clusters were best predicted by alkalinity, watershed area, distance to mainstem, and latitude. This model was the only one that ranked in the top 33% of all possible unaccounted models for both accuracy and precision, and had 3 to 7% higher accuracy than competing models with slightly higher precision. Both the covariate and unaccounted models produced more precise O/E predictions at reference sites (SD = 0.31 and 0.34, respectively) than did their respective null models (both SD = 0.38). The selection process for the censored model included alkalinity, watershed area, distance to mainstem, latitude, and elevation as the best predictors. This model was ranked the highest in precision, was in the top 33% for accuracy, and had 2% higher accuracy than the model with the 2nd highest precision. The censored model also had higher precision than its null model (SD = 0.28 and 0.39, respectively).

Plots of general model performance based on O vs E richness indicate moderate correlation between O and E for all 3 model approaches ($R^2 = 0.50$–0.65). Visual inspection of plots indicated that models frequently over-predicted or under-predicted E, especially as species numbers increased (Fig. 2A–C). However, the intercept and slope comparisons indicated that the models were not biased (range = -0.40–0.20 and 0.90–1.12, respectively; i.e., no regressions differed significantly from a 1 : 1 line). AUC values across all 3 O/E

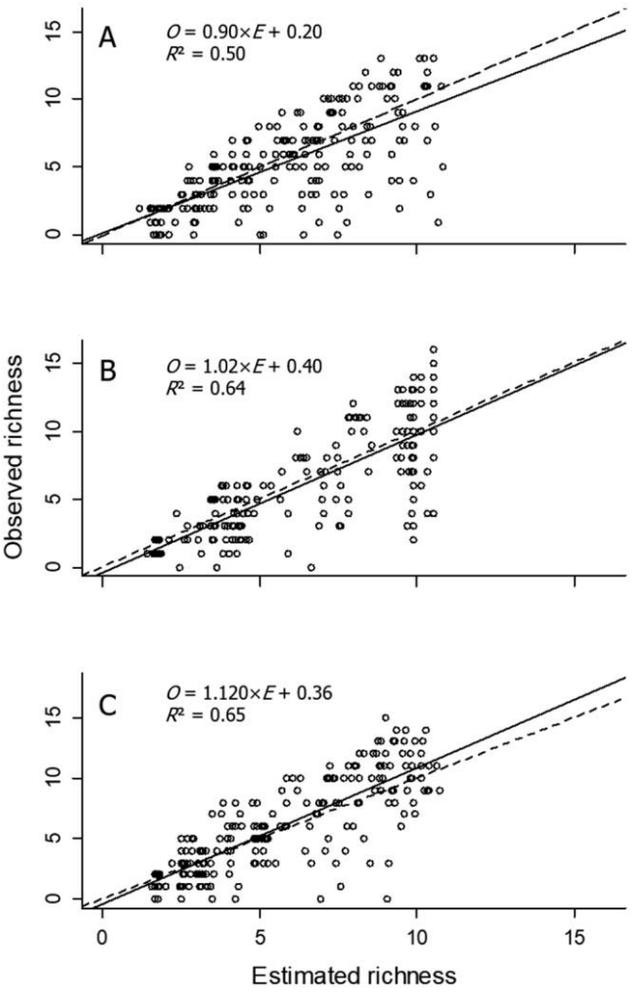

Figure 2. Observed (O) vs expected (E) richness plots for all 215 sample sites as predicted from the covariate (A), censored (B), and unaccounted models (C). The solid line is the simple linear model regression line, and the dashed line is a 1:1 line.

models ranged from 0.52 to 0.84. Nineteen of 21 native species had AUC values >0.7 for the covariate model, as did 14 species for the censored model and 15 species for the unaccounted model (see Table S1, Fig. S4). The covariate model clearly had good performance for more species, but AUC values of the 3 approaches were not significantly different ($F_{2, 62} = 1.00$, $p = 0.374$).

The ability of O/E indices to distinguish between land-use classes varied depending on how models accounted for large Brown Trout at reference sites. The index for the covariate model indicated that all reference conditions (sites with Brown Trout present and absent) had higher taxonomic completeness compared with altered and medium-impact conditions ($F_{3, 211} = 13.06$, $p < 0.001$; Fig. 3A). The censored index was also able to distinguish reference conditions without Brown Trout from altered conditions ($F_{3, 211} = 4.96$, $p = 0.002$; Fig. 3B), and the censored index further indicated that reference sites with Brown Trout had statistically similar O/E values as altered sites (means = 0.85 and 0.87, respectively). Finally, the index for the unaccounted model was unable to distinguish reference conditions from altered conditions ($F_{3, 211} = 2.34$, $p = 0.074$; Fig. 3C).

There were striking differences in the $O_F : E_F$ ratios for several native species across the 3 models. The covariate model identified Brook Trout (*Salvelinus fontinalis*), Common Shiner (*Luxilus cornutus*), Northern Hogsucker (*Hypentelium nigricans*), Rainbow Darter (*Etheostoma caeruleum*), Striped Shiner (*L. chrysocephalus*), and Yellow Bullhead (*Ameiurus natalis*) as sensitive species that decrease with alteration (Table 2). In contrast, only Brook Trout and Yellow Bullhead were identified as sensitive to alteration in the censored and unaccounted models. Hence, changes in the frequency of species occurrence between reference and altered sites appear to be associated more with the specific approach for handling Brown Trout reference sites rather than actual changes in reference conditions.

With respect to index performance, the index for the covariate model that directly accounted for the presence of LBT had the highest responsiveness (Fig. 3A; Table 3). All 3 indices had a weak, positive bias associated with stream size (i.e., higher index scores with increasing size). The covariate index also had the highest sensitivity, identifying 40.5% of altered sites as being below the 10th percentile of reference site scores as compared with 27.0% and 29.7% for the censored and unaccounted indices, respectively. Precision was slightly higher for the indices from the covariate and censored models that accounted for Brown Trout compared with the unaccounted index.

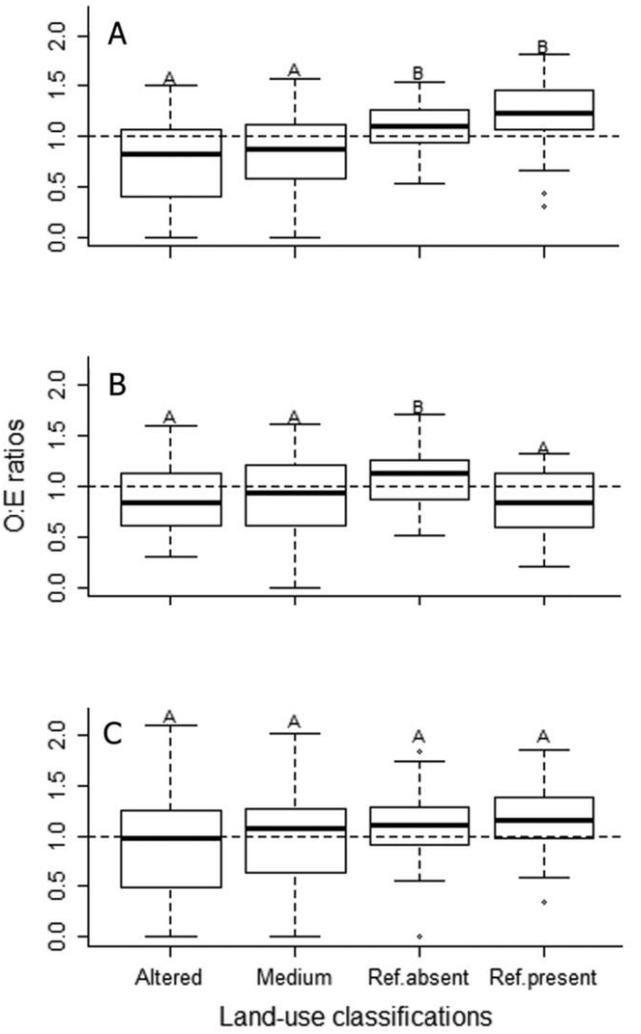

Figure 3. Comparison of O/E ratios for the different land-use classes with expected number of taxa (E) predicted by the covariate (A), censored (B), and unaccounted models (C). An O/E value of 1 (dashed lines) is where the number of observed taxa (O) and E are equal. Upper-case letters within each panel distinguish statistically significant classes for each index based on Tukey's pairwise comparisons ($p < 0.05$). Ref.absent indicates reference sites without large Brown Trout (LBT; $n = 57$) and Ref.present indicates reference sites with LBT ($n = 31$).

DISCUSSION

An extensive literature documents the negative impacts of non-native species on native assemblages (Mitchell and Knouft 2009, Gozlan et al. 2010, Ricciardi et al. 2013), but no studies have explicitly tested either the effects of non-native species on bioassessment indices or potential strategies to account for their impacts. The ability of bioassessments to accurately detect the effects of human-associated stressors on aquatic assemblages depends on the ability to control for background ecological processes that underlie variation in assemblage composition (Allan 2004, Hawkins 2006, Wang et al. 2008, Hawkins et al. 2010). It is well established that naturally-occurring variation in environmental conditions can complicate the ability to detect assemblage changes from physical–chemical alterations, but no one has assessed the influence of biological stressors on the

Table 2. Comparison of $O_F : E_F$ ratios for fish species calculated from probabilities of capture (Pc) derived from the covariate, unaccounted, and censored models. Bolded values identify species as either sensitive (ratios ≤ 0.75; i.e., decreasing occurrence with alteration) or tolerant (ratios ≥ 1.25; i.e., increasing occurrence with alteration) for each respective model. Non-bolded values indicate when a species was not identified as being sensitive or tolerant for a given model.

Group	Species	Covariate model	Unaccounted model	Censored model
Sensitive	Brook Trout	**0.22**	**0.55**	**0.42**
	Common Shiner	**0.60**	0.81	0.82
	Northern Hogsucker	**0.75**	1.09	1.02
	Rainbow Darter	**0.74**	0.95	1.04
	Striped Shiner	**0.50**	0.80	0.80
	Yellow Bullhead	**0.70**	**0.57**	0.80
Tolerant	Fantail Darter	**1.28**	1.13	**1.46**
	Greenside Darter	1.04	1.08	**1.49**
	Longnose Dace	0.82	1.18	**1.51**
	Pumpkinseed	**1.26**	**1.39**	**1.57**

ability of bioassessment to detect those changes. Non-native species have been considered as a potential response metric in MMIs (Angermeier et al. 2000, Wang et al. 2008), and previous taxonomic completeness studies simply acknowledge that accounting for non-native species could enhance the ability of these indices to detect land-use alteration (Kennard et al. 2005, Carlisle et al. 2008, Meador and Carlisle 2009, Mazor et al. 2016).

We demonstrated that overlooking the potential impacts from non-native species could have profound consequences for interpretations of biological integrity (e.g., an inability to identify biologically altered conditions; Fig. 3C). There are 2 general ways in which non-native species could affect bioassessment inferences. The 1st, and perhaps most common, is that the presence of non-native species will be positively correlated with anthropogenic alteration. Non-natives often invade native assemblages occurring within altered streams (Ross 1991, Kennard et al. 2005, Scoppettone et al. 2005, Leprieur et al. 2008, Hermoso et al. 2011). In this context, the impacts of non-native species should be coincident with the impacts of other anthropogenic stressors on fish assemblages at altered sites. The 2nd scenario, which we address in this study, occurs when non-native species occur in streams that are relatively unaltered by physical–chemical stressors. This scenario can compromise bioassessment if non-native species reduce native species richness in reference streams, thus, generating a potential mismeasure of biotic integrity (Scott and Helfman 1999, Kirk et al. 2017). In this study, we demonstrated that TCIs were unable to effectively distinguish reference and altered conditions without accounting for the negative impacts from the non-native species. Hence, it is difficult to determine what constitutes true reference conditions if reference sites include environmental conditions where non-native species are likely to occur (Chessman and Royal 2004). Many bioassessment studies classify reference conditions based on a threshold for non-natural land use (agricultural or urbanization) at the landscape scale (Hawkins et al. 2000, 2010, Carlisle et al. 2008,

Table 3. Performance metrics for the 3 modeling approaches with reference site sample sizes in parentheses. Bolded values indicate the best score for each performance metric across all methods. Bias variables are the variables used in the original models still associated with O/E values. Bias is the amount of variation (R^2) in index scores explained by those predictor variables. Precision is the standard deviation (SD) of reference site scores for each index. Sensitivity is the proportion of altered sites that had O/E values below the 10th percentile of reference site values. The *F*-test statistic quantifies responsiveness—i.e., the magnitude of difference in bioassessment scores among different land-use classes from the analysis of variance (ANOVA). *Indicates the relationship was significant ($p < 0.05$).

Bioassessment method	Bias variables ($p < 0.05$)	Bias (R^2)	Precision (SD)	Sensitivity (%)	*F*-test statistic
Covariate model ($n = 88$)	Area	0.06	0.31	**40.5**	**13.1***
Unaccounted model ($n = 88$)	Area	0.06	0.34	29.7	2.3
Censored model ($n = 57$)	Area	**0.02**	**0.28**	27.0	5.0*

Vander Laan and Hawkins 2014, Mazor et al. 2016), so understanding the impacts of additional stressors is imperative for improving the ability of bioassessments to infer biological integrity.

In our study system, we were faced with the dilemma of accounting for non-native species with known negative impacts at reference sites (see Kirk et al. 2017). Thus, we considered 3 potential approaches for how to handle non-native species within a bioassessment framework. The 1st approach was to exclude sites that had the non-native species present from the reference pool (i.e., censored model; Kennard et al. 2005, Ode et al. 2016). From a bioassessment perspective, censoring reference sites with non-native species seems most appropriate because the entire basis for selecting reference sites is that they are unaffected by anthropogenic stressors. However, from a practical perspective, true reference sites with no impacts from stressors can be rare, and 1 or more undetected stressors could be acting on assemblages (Hawkins et al. 2000, Chessman and Royal 2004, Carlisle et al. 2008). Many species of non-native trout have been introduced around the world (Fausch 2008, McIntosh et al. 2011), often in high-quality habitats with low levels of degradation that would otherwise meet reference criteria for a given region (Wang et al. 2008).

An adequate sample size of reference sites is critical for both calibrating and validating index performance (Van Sickle et al. 2006, Carlisle et al. 2008). If most potential reference sites (i.e., based on landscape and chemistry variables) have a non-native species present, index construction, and likely, index performance, would be compromised if those sites with non-native species were eliminated. Moreover, non-native species impacts often vary across environmental gradients (Ricciardi et al. 2013, Iacarella et al. 2015), including in our study region where community-level impacts of Brown Trout are most prominent in high-quality, medium-sized streams (10–50 km^2) because this species is uncommon in headwater streams (<10 km^2; Kirk et al. 2017, 2018). Thus, censoring reference sites based on the occurrence of non-native species could be unwarranted without knowledge on the extent of their impacts across environmental gradients and without considering how censoring sites might bias the reference pool.

Censoring Brown Trout sites from the reference pool provided adequate sample sizes to distinguish between reference and altered conditions in our study (Fig. 3B), but it also created an inherent bias within our reference dataset. By censoring Brown Trout reference sites, we compromised representation of large streams from our reference pool (Table 1; average watershed area of the 31 censored sites = 17.5 km^2; see also Ode et al. 2016). This reduction in environmental representation did not influence the responsiveness of the censored index, but it did affect species $O_F : E_F$ ratios and the perceived sensitivity of fish species to altered conditions. Greenside Darter (*Etheostoma blennioides*) and Longnose Dace (*Rhinichthys cataractae*) are relatively sensitive species from a bioassessment perspective (sensu Barbour et al. 1999), but our censored model identified them as tolerant species that increase with alteration (Table 2). These 2 native species are rarely found in headwater streams (<10 km^2; Stauffer et al. 2016, Kirk et al. 2017), so our results are probably an artifact of the dataset having greater representation of larger streams at non-reference sites compared with reference sites for the censored model.

The 2nd modeling approach accounted for the non-native species directly within the modeling framework (covariate model). This method is similar to controlling for environmental covariates. The covariate model identified several species (Common and Striped Shiner, Northern Hogsucker, and Rainbow Darter) as sensitive to alteration, which is consistent with their previously documented sensitivity to biological degradation (Barbour et al. 1999). These species were not identified as sensitive in either the censored or the unaccounted models (Table 2), suggesting the changes in $O_F : E_F$ ratios among approaches are more associated with a species' sensitivity to stream-size or susceptibility to LBT impacts (see Kirk et al. 2017). These results provide further examples of how censoring sites from the reference pool and not accounting for non-native species impacts can lead to incorrect inferences about species responses and perceived impacts of land-use alteration (e.g., Ode et al. 2016). The downside to directly accounting for non-native species impacts is that this approach could generate a perception that reference sites with non-native Brown Trout have high biological integrity. In reality, however, those sites are probably taxonomically incomplete because Brown Trout can reduce or eliminate a subset of native species that would otherwise occur at reference sites (cf. Fig. 3A and 3B; see also Kirk et al. 2017).

The 3rd approach was to ignore any potential Brown Trout impacts and treat them as a potential part of the modeled assemblage (unaccounted model). Bioassessment predominately focuses on native species integrity, but incorporating non-native species in these indices could improve detection of alteration or achieve additional management objectives (Carlisle et al. 2008, Meador and Carlisle 2009). Despite Brown Trout being an invasive species with pervasive impacts on native fish, this species also has high recreational and commercial value as a gamefish (Budy and Gaeta 2017). Thus, managers might want to account for the socioeconomic benefits of this or other gamefish species in their assessment framework and treat them as part of a reference assemblage that managers want to achieve. However, the covariate model still performed better than the unaccounted model in nearly all performance measures when quantifying native integrity (Table 3). Hence, accounting for non-native species impacts is better for detecting effects of physical–chemical degradation rather than ignoring whether species are non-native altogether.

The censored and covariate indices were both successful in distinguishing reference and altered conditions, but comparisons of index performance between the two was still confounded by differences in environmental characteristics among the 2 reference datasets used for prediction (Table 1). For example, the censored index required less study on environmental variation, which could attribute to higher precision and lower bias results compared to other models (Table 3). Our indices also have low precision (SD = 0.28–0.34) compared with indices developed for actual bioassessment programs (SD < 0.20; from Vander Laan and Hawkins 2014). The low precision in our study may have been caused by several factors including small reference site sample sizes, high similarity among fish assemblages associated with a limited stream-size gradient (i.e., $\leq$50 km^2; Kirk et al. 2017), or a small number of taxa being modeled (n = 21; Meador and Carlisle 2009). We acknowledge that it might be possible to improve index performance for these streams, but our study provides a model example of why non-native species impacts need to be accounted for and how bioassessment practitioners can accomplish that in different contexts.

Ultimately, deciding on the most suitable approach for bioassessment practitioners will depend upon the severity of non-native species impacts, data limitations, and management objectives (Hawkins 2006, Mazor et al. 2016). Large-bodied piscivores are likely to be among the most important non-native species to account for in bioassessments because they often have more pronounced effects on native species composition compared with non-natives at lower trophic positions (Mitchell and Knouft 2009). If reference sites are abundant and non-native species occurrences represent a small proportion of total reference sites, censoring sites should not compromise the existing reference pool (Kennard et al. 2005). However, if high-quality (i.e., pristine) reference sites are rare and non-native occurrences represent a large proportion of reference sites, such that censoring sites would result in an incomplete representation of environmental conditions (Ode et al. 2016; Table 1), then directly accounting for non-native species within modeling frameworks should be the most suitable approach (covariate model).

These approaches can also be used for different management objectives, depending on whether bioassessment practitioners want to directly quantify the impacts of these stressors or produce indices that are insensitive to their confounding effects. If the management objective is to identify whether non-native species have a negative impact on biotic integrity under what are otherwise reference conditions, then non-native sites should be censored and analyzed independently as a set of non-reference sites (e.g., Fig. 3B). If the management objective is to infer non-point stressor impacts from land-use change and associated impacts at larger scales, which is the most frequent application of bioassessment studies (Carlisle et al. 2008), then directly accounting for the confounding effects of non-native species would be a more suitable approach. Finally, practitioners could evaluate scores from multiple indices to achieve a more holistic interpretation of biological integrity and decide which approaches produce results most relevant to management goals.

Conclusions

Our study provides a framework for understanding how non-native species can confound the development and interpretation of bioassessment tools. O/E TCIs based on fish species occurrence could distinguish reference and altered conditions only when non-native Brown Trout were accounted for in the covariate model or when reference sites with Brown Trout were censored from the reference pool. Incorporating Brown Trout reference sites without accounting for their presence led to an overestimation of taxonomic completeness at altered sites and, thus, biased assessments of biological integrity because these non-native, top predators reduce the native species richness in reference streams. Systems that have highly invasive, widespread species with well-documented impacts on native assemblages might represent the best examples for further addressing how non-native species can alter interpretations of biological integrity. Our results have important implications for bioassessment studies because reference conditions are typically derived from landscape-level criteria and potential impacts from the presence of non-native species are often overlooked. Our results emphasize the need to disentangle the effects of non-native, invasive predators from environmental gradients and other human stressors to better explain the multiple underlying causes of variation in assemblage composition. Our study provides bioassessment practitioners with 3 modeling approaches for accounting for the impacts of non-native species that can be chosen based on management objectives. Understanding how non-native species affect measures of biological integrity should ultimately improve our ability to use bioassessments in aquatic ecosystems by providing better insight into what constitutes the most appropriate characterization of reference conditions (Chessman and Royal 2004, Ode et al. 2016).

ACKNOWLEDGEMENTS

Author contributions: MAK and SAW collected the fish assemblage data and conceived the manuscript. MAK analyzed data and wrote the paper. MAK and SAW critically revised and edited the manuscript.

SAW received funding from Allegheny College and grants via the National Fish and Wildlife Federation and the Pennsylvania Fish and Boat Commission through the Unassessed Waters Initiative. We thank Brandon Goeller, Leslie Rieck, Lynette Gardner, Ashley Seitz, Elizabeth Goetz, Kirsten Ressel, Anna Zimmerman,

Katie Robbins, Scott Kirk, Jared Balik, Susan Washko, Maggie McClain, Anna Rosswog, Logan Billet, and Kaitlyn Campbell, who all contributed to data collection and analysis in early years of this study. We thank Chris Shaffer for GIS support. John Van Sickle graciously provided code and statistical guidance that assisted with the development of taxonomic completeness models. Finally, we thank Brandon Goeller, Bryan Maitland, Peter Ode, Lester Yuan, Raphael Mazor, Chuck Hawkins, and 3 anonymous reviewers who provided constructive comments that improved the quality of the manuscript.

LITERATURE CITED

Adler, R. W. 2012. The decline and (possible) renewal in aspiration in the Clean Water Act. Washington Law Review 88:759–812.

Allan, J. D. 2004. Landscape and riverscapes: The influence of land use on stream ecosystems. Annual Review of Ecology, Evolution and Systematics 35:257–284.

Angermeier, P. L., and J. R. Karr. 1994. Biological integrity versus biological diversity as policy directives. Bioscience 44:690–697.

Angermeier, P. L., R. A. Smogor, and J. R. Stauffer. 2000. Regional frameworks and candidate metrics for assessing biotic integrity in Mid-Atlantic highland streams. Transactions of the American Fisheries Society 129:962–981.

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: Periphyton, benthic macroinvertebrates and fish. 2nd edition. EPA 841-B-99-002. United States Environmental Protection Agency, Office of Water, Washington DC.

Budy, P., and J. W. Gaeta. 2017. Brown trout as an invader: A synthesis of problems and perspectives in North American. Pages 525–544 *in* J. Lobón-Cerviá and N. Sanz (editors). Brown trout: Biology, ecology, and management. Wiley, Hoboken, New Jersey.

Canning, A. D. 2018. Predicting New Zealand riverine fish reference assemblages. PeerJ 6:e4890.

Carlisle, D. M., C. P. Hawkins, M. R. Meador, M. Potapova, and J. Falcone. 2008. Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages. Journal of the North American Benthological Society 27:116–37.

Chessman, B. C., and M. J. Royal. 2004. Bioassessment without reference sites: Use of environmental filters to predict natural assemblages of river macroinvertebrates. Journal of North American Benthological Society 23:599–615.

Fausch, K. D. 2008. A paradox of trout invasions in North America. Biological Invasions 10:685–701.

Gozlan, R. E., J. R. Britton, I. Cowx, and G. H. Copp. 2010. Current knowledge on non-native freshwater fish introductions. Journal of Fish Biology 76:751–786.

Hawkins, C. P. 2006. Quantifying biological integrity by taxonomic completeness: Its utility in regional and global assessments. Ecological Applications 16:1277–1294.

Hawkins, C. P., Y. Cao, and B. Roper. 2010. Method of predicting reference condition biota affects the performance and interpretation of ecological indices. Freshwater Biology 55:1066–1085.

Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10:1456–1477.

Hermoso, V., M. Clavero, F. Blanco-Garrido, and J. Prenda. 2011. Invasive species and habitat degradation in Iberian streams: An analysis of their role in freshwater fish diversity loss. Ecological Applications 21:175–188.

Hill, R. A., E. W. Fox, S. G. Leibowitz, A. R. Olsen, D. J. Thornbrugh, and M. H. Weber. 2017. Predictive mapping of the biotic condition of conterminous U.S. rivers and streams. Ecological Applications 27:2397–2415.

Iacarella, J., J. T. A. Dick, M. E. Alexander, and A. Ricciardi. 2015. Ecological impacts of invasive alien species along temperature gradients: Testing the role of environmental matching. Ecological Applications 25:706–716.

Karr, J. R. 1991. Biological Integrity: A long-neglected aspect of water resource management. Ecological Applications 1:66–84.

Kennard, M. J., A. H. Arthington, B. J. Pusey, and B. D. Harch. 2005. Are alien fish a reliable indicator of river health? Freshwater Biology 50:174–193.

Kirk, M. A., A. N. Rosswog, K. N. Ressel, and S. A. Wissinger. 2018. Evaluating the trade-offs between invasion and isolation for native brook trout and non-native brown trout in Pennsylvania streams. Transactions of the American Fisheries Society 147:806–817.

Kirk, M. A., S. A. Wissinger, B. C. Goeller, and L. O. Rieck. 2017. Co-varying impacts of land use and non-native brown trout on fish communities in small streams. Freshwater Biology 62: 600–614.

Kocovsky, P. M., and R. F. Carline. 2005. Stream pH as an abiotic gradient influencing distributions of trout in Pennsylvania streams. Transactions of the American Fisheries Society 134: 1299–1312.

Leprieur, F., O. Beauchard, S. Blanchet, T. Oberdorff, and S. Brosse. 2008. Fish invasions in the world's river systems: When natural processes are blurred by human activities. PLoS Biology 6(2):e28.

Mazor, R. D., A. C. Rehn, P. R. Ode, M. Engeln, K. C. Schiff, E. D. Stein, D. J. Gillett, D. B. Herbst, and C. P. Hawkins. 2016. Bioassessment in complex environments: Designing an index for consistent meaning in different settings. Freshwater Science 35:249–271.

McDowall, R. M. 2003. Impacts of introduced salmonids on native galaxiids in New Zealand upland streams: A new look at an old problem. Transactions of the American Fisheries Society 132:229–238.

McIntosh, A. R. 2000. Habitat and size-related variations in exotic trout impacts on native galaxiid fishes in New Zealand streams. Canadian Journal of Fisheries and Aquatic Sciences 57:2140–2151.

McIntosh, A. R., P. A. McHugh, and P. Budy. 2011. Brown trout (*Salmo trutta*). Pages 285–296 *in* R.A. Francis (editor). A handbook of global freshwater invasive species. Earthscan, London, UK.

Meador, M. R., and D. M. Carlisle. 2009. Predictive models for fish assemblages in eastern U.S. streams: Implications for assessing biodiversity. Transactions of the American Fisheries Society 138:725–740.

Mitchell, A. L., and J. H. Knouft. 2009. Non-native fishes and native species diversity in freshwater fish assemblages across the United States. Biological Invasions 11:1441–1450.

Ode, P. R., A. C. Rehn, R. D. Mazor, K. C. Schiff, E. D. Stein, J. T. May, L. R. Brown, D. B. Herbst, D. Gillett, K. Lunde, and C. P. Hawkins. 2016. Evaluating the adequacy of a reference site pool for ecological assessments in environmentally complex regions. Freshwater Science 35:237–248.

Piñeiro, G., S. Perelman, J. P. Guerschman, and J. M. Paruelo. 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? Ecological Modelling 216:316–322.

Rahel, F. J. 2000. Homogenization of fish faunas across the United States. Science 288:854–856.

Ricciardi, A., M. F. Hoopes, M. P. Marchetti, and J. L. Lockwood. 2013. Progress towards understanding the ecological impacts of nonnative species. Ecological Monographs 83:263–282.

Rose, P. M., M. J. Kennard, D. B. Moffatt, F. Sheldon, and G. L. Butler. 2016. Testing three species distribution modelling strategies to define fish assemblage reference conditions for stream bioassessment and related applications. PLoS ONE 11(1):e0146728.

Ross, S. T. 1991. Mechanisms structuring stream fish assemblages: Are there lessons from introduced species? Environmental Biology of Fishes 30:359–368.

Ruaro, R., and E. A. Gubiani. 2013. A scientometric assessment of 30 years of the index of biotic integrity in aquatic ecosystems: Applications and main flaws. Ecological Indicators 29:105–110.

Sánchez-Montoya, M. M., M. R. Vidal-Abarca, T. Puntí, J. M. Poquet, N. Prat, M. Rieradevall, J. Alba-Tercedor, C. Zamora-Muñoz, M. Toro, S. Robles, and M. Álvarez. 2009. Defining criteria to select reference sites in Mediterranean streams. Hydrobiologia 619:39–54.

Scoppettone, G. G., P. H. Rissler, C. Gourley, and C. Martinez. 2005. Habitat restoration as a means of controlling non-native fish in a Mojave Desert oasis. Restoration Ecology 13:247–256.

Scott, M. C., and G. S. Helfman. 1999. Native invasions, homogenization, and mismeasure of integrity of fish assemblages. Fisheries 26:6–15.

Simon, K. S., and C. R. Townsend. 2003. Impacts of freshwater invaders at different levels of ecological organization, with emphasis on salmonids and ecosystem consequences. Freshwater Biology 48:982–994.

Southerland, M. T., G. M. Rogers, M. J. Kline, R. P. Morgan, D. M. Boward, P. F. Kazyak, R. J. Klauda, and S. A. Stranko. 2007. Improving biological indicators to better assess the condition of streams. Ecological Indicators 7:751–767.

Sowersby, W., R. M. Thompson, and B. B. M. Wong. 2015. Invasive predator influences preferences in a freshwater fish. Environmental Biology of Fishes 99:187–193.

Stauffer, J. R., R. W. Criswell, and D. P. Fischer. 2016. The fishes of Pennsylvania. Cichlid Press, El Paso, Texas.

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. Setting expectations for the ecological condition of streams: The concept of reference condition. Ecological Applications 16:1267–1276.

Strayer, D. L. 2010. Alien species in fresh waters: Ecological effects, interactions with other stressors, and prospects for the future. Freshwater Biology 55:152–174.

Turek, K. C., M. A. Pegg, K. L. Pope, and S. Schainost. 2016. Potential population and assemblage influences of non-native trout on native nongame fish in Nebraska headwater streams. Ecology of Freshwater Fish 25:99–108.

Van Sickle, J., D. D. Huff, and C. P. Hawkins. 2006. Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. Freshwater Biology 51: 359–372.

Vander Laan, J. J., and C. P. Hawkins. 2014. Enhancing the performance and interpretation of freshwater biological indices: An application in arid zone streams. Ecological Indicators 36:470–482.

Wagner, T., J. T. DeWeber, J. Detar, and J. A. Sweka. 2013. Landscape-scale evaluation of asymmetric interactions between brown trout and brook trout using two-species occupancy models. Transactions of the American Fisheries Society 142:353–361.

Wang, L., T. Brenden, P. Seelbach, A. Cooper, D. Allan, R. Clark, and M. Wiley. 2008. Landscape based identification of human disturbance gradients and reference conditions for Michigan streams. Environmental Monitoring and Assessment 141:1–17.

Wright, J. F. 1995. Development and use of a system for predicting the macroinvertebrate fauna in flow waters. Australian Journal of Ecology 20:181–197.

Zimmerman, J. K. H., and B. Vondracek. 2006. Interactions of slimy sculpin (*Cottus cognatus*) with native and nonnative trout: Consequences for growth. Canadian Journal of Fisheries and Aquatic Science 63:1526–1535.