A needle in the haystack? Applying species co-occurrence frameworks with fish assemblage data to identify species associations and sharpen ecological hypotheses

Zachery D. Zbinden (Zach Zbinden; ZDZ)

Department of Biological Sciences

850 W. Dickson St., Room 601,

University of Arkansas, Fayetteville, Arkansas 72702

E-mail: zdzbinde@uark.edu

Accepted Article

Abstract

Different species can associate or interact in many ways, and methods exist for inferring associations and underlying mechanisms from incidence data (e.g., co-occurrence frameworks). These methods have received much criticism despite their recent resurgence in the literature. However, co-occurrence frameworks for identifying non-randomly associated species pairs (e.g. aggregated or segregated pairs) have value as heuristic tools for sharpening hypotheses concerning fish ecology. Here, I aimed to provide a case study examining species co-occurrence across 33 stream fish assemblages in southeastern Oklahoma, USA, which were sampled twice (1974 & 2014). I sought to determine: 1) which species were non-randomly associated, 2) what processes might have driven these associations, and 3) how consistent were patterns across time? Associations among most pairs of species (24 species; 276 unique pairs) were not significantly different from random (>80%). Among all significant, non-randomly associated species pairs (54 unique pairs), 78% (42 pairs) were aggregated and 22% (12 pairs) segregated. Most of these (28 pairs, 52%) were hypothesized to be driven by non-biotic mechanisms: habitat filtering (20 pairs; 37%), dispersal limitation (2 pairs; 0.4%), or both (6 pairs; 11%). The remaining 26 non-randomly associated pairs (48%) had no detectable signal of spatial or environmental factors involved with the association; therefore, the potential for biotic interaction was not refuted. Only 5 species pairs were consistently associated across both sampling periods: stonerollers *Campostoma spp.* and orangebelly darter *Etheostoma radiosum;* red shiner *Cyprinella lutrensis* and bullhead minnow *Pimephales vigilax*; bluegill sunfish *Lepomis macrochirus* and redear sunfish *Lepomis microlophus*; redfin shiner *Lythrurus umbratilis* and bluntnose minnow *Pimephales notatus*; and bigeye shiner *Notropis boops* and golden shiner *Notemigonus crysoleucas*. Frameworks for identifying non-randomly associated species pairs can provide

insight into broader mechanisms of species assembly and point to potentially interesting species interactions (out of many possible pairs). However, this approach is best applied as a tool for sharpening hypotheses to be investigated further. Rather than a weakness, the heuristic nature is the strength of such methods, and can help guide biologists toward better questions by employing relatively cheap diversity survey data, that are often already in hand, so as to reduce complex interaction networks down to their non-stochastic parts which warrant further investigation.

1 | Introduction

Interactions among species are an important piece of the ecological and evolutionary puzzle (Benton, 2009; Schemske, Mittelbach, Cornell, Sobel, & Roy, 2009), and, because of inter-dependence among species, a key component of wholistic conservation (Gilman, Urban, Tewksbury, Gilchrist, & Holt, 2010; Wisz *et al*., 2013). Within and among assemblages there are a myriad of ways individuals of different species can interact in space and time. Considerable attention has been given to competition (Bonin, Boström, Boström-Einarsson, Munday, & Jones, 2015; Schluter, 1994) which has waxed and waned considerably through time (Matthews, 1998; Wiens, 1977). Other examples include predator/prey (or parasite/host) interactions (Aragón & Sánchez-Fernández, 2013; Power, Matthews, & Stewart, 1985; Zurell, Pollock, & Thuiller, 2018) and less commonly studied facilitative and mutualistic interactions (Peoples & Frimpong, 2016; Silknetter et al., 2020; Stachowicz, 2001). Even in the absence of biotic interaction

among species, other forces can lead to certain species co-occurring together more or less often than expected by chance alone (aggregation or segregation, respectively), including: habitat filtering, niche partitioning, and species packing (Geheber & Frenette, 2016; Heino & Grönroos, 2013; Earl E. Werner, 1977), spatial filtering via dispersal limitation or historical contingency (Blois *et al.*, 2014; Gilpin & Diamond, 1982; Schamp *et al.*, 2015), and neutral or stochastic dynamics such as ecological drift (Bell, 2005; Hubbell, 2001; Ulrich, 2004). Further adding to the complexity of the inter-dependent web among interacting species is the element of time and space. For example, changes in density of one or both interactors through time due to predation or disturbance (McMeans *et al.*, 2020; Wiens, 1977) and/or seasonality (Brimacombe, Bodner, & Fortin, 2020), or through space due to environmental gradients (Case & Taper, 2000; Schwilk & Ackerly, 2005), can obscure—or at least alter—any pattern of association we might expect to observe.

Detecting species interactions has often relied on controlled experiments or direct natural history observation (Silknetter et al., 2020). Even in small systems the number of interactions increases quickly with number of species until it becomes intractable to assess each individually (Morales-Castilla, Matias, Gravel, & Araújo, 2015). Therefore, ecologists have developed methods to statistically test for non-random associations among species from incidence data as a means to circumvent such an arduous task (Forbes, 1907; Pielou & Pielou, 1968). These methods were largely born out of an interest in identifying competition among species (Diamond, 1975), but were later generalized using null modelling to provide more rigorous tests for identifying patterns of both aggregation and segregation (Connor & Simberloff, 1979; Gotelli, 2000; Peres-Neto, 2004). Because of the multitude of factors, both biotic and abiotic, which can drive

patterns of association (Barner *et al.*, 2018; Blanchet *et al.*, 2020), more recent work has combined earlier methods for identifying non-random associations with additional multivariate approaches (e.g., Blois *et al.*, 2014) which incorporate spatial and environmental data to disentangle the potential ecological mechanisms driving observed patterns (Cordero & Jackson, 2019; Devercelli, Scarabotti, Mayora, Schneider, & Giri, 2016; Sfenthourakis, Tzanatos, & Giokas, 2006).

Both methods for identifying associations and those more recent extensions aimed at parsing them have faced considerable scrutiny (Blanchet et al., 2020; Peterson, Soberón, Ramsey, & Osorio-Olvera, 2020; Thurman, Barner, Garcia, & Chestnut, 2019). Methods for identifying associations face several issues. For one, as with many ecological methods, there are a variety of approaches and they are not all equally robust for detecting associations (Barner *et al.*, 2018; Lavender, Schamp, Arnott, & Rusak, 2019). Furthermore, ecological interactions may not lead to the clear spatial patterns that we expect (Freilich, Wieters, Broitman, Marquet, & Navarrete, 2018) and thus these methods can fall short of detecting them (Cazelles, Araújo, Mouquet, & Gravel, 2016). In addition, and as with all things ecological, sampling biases caused by the difficult detection of rare species (Calatayud *et al.*, 2020; MacKenzie, Bailey, & Nichols, 2004) and effects of spatial scale of sampling grain (Perry, Miller, Lamont, & Enright, 2020) can affect the ability of these methods to reliably identify real associations (Peterson et al., 2020). As for interpreting ecological mechanisms from co-occurrence data, it can be difficult to determine whether the pattern is driven by habitat filtering or biotic processes because it is impossible to account for each environmental facet that may govern an association, therefore unmeasured factors confound such conclusions (Blanchet *et al.*, 2020). Also interaction networks (more than

two species) preclude pairwise analyses from detecting effects of additional species which may have direct or indirect effects on either member of a pair (Cazelles *et al*., 2016; Holt & Bonsall, 2017). Other properties of ecosystems such as habitat gradients, demographic dynamics, and temporal changes (Brimacombe et al., 2020) all play a role in shifting species associations and our ability to detect them, especially with incidence data that merely provide a "snapshot" of diversity in space and time (Blanchet *et al*., 2020).

Species incidence is among the easiest of ecological data to gather, and so it is reasonable to hope to extract as much information as possible (e.g., species associations) from such data (Blanchet *et al*., 2020). One of the biggest issues with these methods is what users expect to get out of them. Because of the issues described above, it is not reasonable to expect a definitive test of species associations and mechanisms from such methods. Rather, species co-occurrence analyses are an exploratory tool (Silknetter et al., 2020) for finding statistically non-random associations among large sets of potentially interacting pairs and sharpening hypotheses regarding the potential underlying mechanisms. These associations should be further evaluated through direct observation or experimental means to test these hypotheses. Given the large number of species interactions in even small communities (e.g., 10 species = 45 possible pairwise associations), these methods can help us find the "needles in the haystack," so to speak, and ask more focused questions about how fish species interact.

Here, I apply two frameworks to investigate freshwater fish species associations and how they have changed between two sampling campaigns. For these tests, I was interested in several questions: 1) which species pairs are non-randomly associated and how many are there? 2) what

do these frameworks tell us about mechanisms driving species associations? And 3) are species associations consistent among two sampling campaigns executed 40 years apart?

2 | Materials and Methods

2.1 / Study region & fish collection

The study region is in the south-central United States (southeastern Oklahoma) within a relatively rugged topographic region, known as the U.S. Interior Highlands (Figure 1). Muddy Boggy Creek and Clear Boggy Creek merge to form the Muddy Boggy River, which is a tributary to the Red River of the South (Mississippi Basin). The Muddy Boggy River flows southward and drains 6,291 km^2 from an area approximately 113 km (north-south) by 48 km (east-west). See Pigg (1977) and Zbinden & Matthews (2017) for more detailed descriptions of the region.

Fish collections were made in the Muddy Boggy River drainage between 1974-1975 (Pigg, 1977), and between 2014-2015 (Zbinden & Matthews, 2017). From these collections, 33 sites sampled at both times were included based on a previous study that used these sites to quantify fish assemblage change over time (Zbinden, 2020). All collections were made between May and September during low flows. Seines of the appropriate size relative to the sampling reach (1.22 m x 2.44 or 4.57 m) with mesh sizes approximately 4.88 mm were used to sample all available habitat within a stream reach transect 100 m in length. The fish collected at a site/locality constitute the "fish assemblage", which is synonymous with the term "fish community" used by others for the sake of consistency (e.g. Matthews & Marsh-Matthews, 2017). However, I choose to use assemblage because fish alone are a phylogenetically restricted piece of the whole

community, therefore an assemblage (*sensu* Fauth, Bernardo, & Camara, 1996). Prior to analyses, only species captured during both sampling periods were retained, and species occurring at <6 sites in both time periods were removed because of low statistical power for testing associations (Cordero & Jackson, 2019; Gotelli & McCabe, 2002; Veech, 2013). Two taxonomic changes were required: central stoneroller *Campostoma anomalum* (Rafinesque 1820) and highland stoneroller *Campostoma spadiceum* (Girard 1856) collected in 2014 were collapsed into *Campostoma spp.* because recognition of *C. spadiceum* did not occur until after 1975 (Cashner, Matthews, Marsh-Matthews, Unmack, & Cashner, 2010) and thus were not differentiated by J. Pigg. In the same manner, blackstripe topminnow *Fundulus notatus* (Rafinesque 1820) and blackspotted topminnow *Fundulus olivaceus* (Storer 1845) were collapsed into *Fundulus spp.* due to identification issues in this region (J.F. Schaefer and W.J. Matthews, pers. comm.)

2.2 / Spatial data

I calculated pairwise fluvial network distance among all fish sampling locations using the Network Analyst Extension in ArcGIS v.10.8 (ESRI). This distance matrix was then decomposed into a rectangular matrix of spatial eigenvectors using distance-based Moran's eigen vector map analysis (dbMEM; Stéphane Dray *et al.*, 2006), also known as principal coordinates of neighbor matrices (Borcard & Legendre, 2002). The spatial eigenvectors were generated using the function `dbmem` in the R statistical software (v.3.6.1, R Core Team, 2019) package `adespatial` (Dray *et al.*, 2020), and I only retained eigenvectors with positive spatial autocorrelation (6 MEMs were retained).

2.3 / Environmental data

Environmental factors for each collection site were characterized using data provided by HydroATLAS v.1.0 at 15 arc-second (~500m) resolution (Linke *et al*., 2019). This global compendium of sub-basin and river reach characteristics contains 56 hydro-environmental factors partitioned into 281 attributes drawn from a multitude of global datasets (Linke et al., 2019). The RiverATLAS v.10 shapefile for North America was trimmed in ArcMAP (ESRI v.10.8) to include only the subset of river network necessary for this study. This shapefile represented the river network and all associated environmental variables along the stream reaches, which are defined as stretches of stream between confluences or confluence and terminus. Geographic coordinate data of sampling sites were imported into the GIS and exported as a shapefile. Both shapefiles were projected using NAD 1983 (2011) Contiguous USA Albers projection. Environmental attributes of the river shapefile were joined to the appropriate sites using R packages `raster` (Hijmans & van Etten, 2015) and `sf` (Pebesma, 2019). From the total set (281 attributes), 22 variables were chosen because I expected they would be similar between sampling periods or because they represented long term averages. To be clear, the environmental data was identical between each sampling period, but the characteristics were chosen specifically to match that assumption.

The environmental variables broadly characterize the hydrologic, physiographic, and geologic nature of the sampling locations. These reach catchment variables (unless otherwise noted) included: Strahler stream order, mean annual discharge (1971-2000), mean annual land surface runoff (1971-2000), maximum inundation extent (1993-2004), river area, river area in the upstream catchment, river volume, river volume in the upstream catchment, mean groundwater

table depth (1927-2009), mean elevation, mean terrain slope, mean stream gradient, mean annual precipitation (1950-2000), extent of deciduous tree cover in reach catchment, extent of evergreen tree cover in reach catchment, extent of herbaceous cover in reach catchment, total forest cover in reach catchment, mean fraction of clay in soil of reach catchment, mean fraction of silt in soil of reach catchment, mean fraction of sand in soil of reach catchment, mean organic carbon fraction in soil of reach catchment, and spatial extent of exposed karst in reach catchment. HydroATLAS (Linke *et al.*, 2019) codes for environmental factors are retained in the data file (Supplementary Material) so that they can be compared to the original database. Descriptions above are based on the HydroATLAS catalog (https://hydrosheds.org/page/hydroatlas).

All environmental factors were standardized by mean and unit standard deviation before being reduced using principal components analysis, using function `prcomp` of the R package `stats` (RCoreTeam, 2019) without scaling or centering because variables had already been standardized. Six of the principal components were retained to make the number of factors equal between spatial and environmental sets (6 PCs and 6 MEMs). These 6 principal components accounted for 87.6% of variation in the environmental set. See Supplementary Material for full environmental PCA details and factor loadings on each component.

I used two testing frameworks (Figure 2) aimed at detecting significant non-random species associations and subsequently hypothesizing mechanisms that govern associations by incorporating the spatial and environmental vectors (Blois *et al.*, 2014; D'Amen, Mod, Gotelli, & Guisan, 2018; Gotelli & Ulrich, 2010; Veech, 2013). The Null Model Framework (Classic Framework *sensu* D'Amen *et al.* 2018) and The Probabilistic Framework (Veech, 2013) differ

only in how significantly non-random species associations are tested, but the same decision tree approach for statistically deducing mechanisms via multivariate analysis of variance is employed (Descriptions below).

2.4 | Null Model Framework (NMF)

The null model framework uses C-score index, a measure of "checkerboardedness" among species pairs (Stone & Roberts, 1990). The original index measured the average C-score for the entire species incidence matrix, but here an index for each species pair is desired. The following equation was used:

$$Cscore_{ab} = \frac{(R_a - S) \times (R_b - S)}{(R_a \times R_b)}$$

Where the number of occurrences of species a is R_a and species b is R_b, S is the number of sites where both species were collected. A maximally aggregated species pair (all occurrences shared) will have a C-score of 0, and a maximally segregated species pair (no sites in common) will have a C-score of 1.

C-score significance for each species pair was tested by generating 10,000 null assemblages by randomly reshuffling the observed matrix. Column sums (species incidences) and row sums (per site species richness) of null assemblages were constrained to be equal to the observed matrix (fixed-fixed randomization *sensu* Connor & Simberloff, 1979; Gotelli, 2000). Observed C-scores for each pair were compared to the appropriate null distribution to test for significance (α = 0.05). Where n is equal to the number of unique species in the presence-absence matrix, the total

number of species pairs tested is equal to $n((n - 1)/2)$. To reduce false detection error rate I used a mean-based Bayes approach (Bayes M criterion) from Gotelli and Ulrich (2010) and originally described by Efron (2005). Significantly associated species pairs are grouped as either aggregated or segregated by comparing observed C-score with the mean randomized score for a given species pair. Z-scores were calculated based on 10,000 randomizations. Large positive scores are associated with segregating species, and large negative scores are associated with aggregating species (D'Amen *et al.*, 2018; Gotelli & Ulrich, 2010). Results from null model analysis were then used with a deductive logic tree approach described below to hypothesize processes underlying each significant association (Figure 2).

2.5 | Probabilistic Framework (PRF)

The Probabilistic Framework (PRF) is an entirely different approach for testing patterns of species co-occurrence based on probability theory. This approach uses combinatorics (i.e., probability within finite sets) to estimate the probability of species co-occurring at greater than or less than observed sites by considering the possible combinations a given pattern of species distributions in *N* sites could possibly be observed (*Eq 1.* and *Fig. 2* from Veech, 2013). These probabilities may be obtained analytically and do not require a null distribution or data randomization which can itself introduce Type I and II error (Lavender *et al.*, 2019; Veech, 2014). The probabilistic approach determines whether species are positively (i.e., aggregated), negatively (i.e., segregated), or randomly associated. The probabilistic calculations are implemented in the R package `cooccur` (Griffith, Veech, & Marsh, 2016). Although the Null Model Framework (NMF) and Probabilistic Framework (PRF) differ in mathematical approach,

both result in a set of significantly associated species pairs which can be further analyzed as described below.

2.6 | Decision tree for statistically deducing mechanisms

Analyses of species pairwise co-occurrence presented above cannot determine processes responsible for non-random species associations by themselves, although they are often inferred as indicative of biotic interaction (Peterson et al., 2020). However, by including additional analyses of spatial and environmental data for each site, it is possible to determine whether dispersal limitation or habitat filtering are likely forces driving non-random associations, rather than biotic interaction. If neither of those forces are significant, then it is not possible to refute a biotic interaction is possible. However, this hypothesis should obviously be tested further through direct observation or experimentation. Both the NMF and PRF incorporate the same decision tree approach for disentangling mechanisms (Figure 2). This approach is modified from Blois *et al.* (2014) and D'Amen *et al.* (2018).

Ultimately, multivariate analysis of variance (MANOVA) of different site categories is used to test for significant spatial and environmental effects on the non-random associations inferred from the analyses described above. For each significantly associated species pair, the sites in the incidence matrix belong to one of four mutually exclusive site-type categories: both species present (1,1), neither species present (0,0), or one species present without the other (1,0) *and* (0,1). Environmental and spatial data are partitioned between site-type categories to test for significant differences among them via MANOVA. These tests inform the decision tree and lead

to logical, deductive inferences regarding potential mechanisms involved in species pair associations (Figure 2).

Allotopic sites, having one species without the other (1,0) and (0,1), are contrasted to discern mechanisms for *segregated* species pairs. If environmental factors and/or spatial arrangement differ among these two site classes, then the pattern of segregation is likely driven by those factors. However, if neither environment nor spatial factors differ, then negative biotic interaction (e.g., competition or predation) is not rejected. However, this case cannot exclude other alternative hypotheses regarding segregation. For example, the pattern could be driven by differences in unmeasured environmental factors. Syntopic sites where both species were found (1,1) and sites with neither species (0,0) are compared to discern potential mechanisms for *aggregated* species pairs in the same manner. Again, in the absence of significant environmental or spatial difference between site-type classes, the possibility of a biotic interaction remains. This could include several interactions including facilitation, predation, or indirect effects with other species (Blois *et al.*, 2014), or unmeasured environmental factors.

Within each of the 4 frameworks (NMF1974, NMF2014, PRF1974, & PRF2014), and following the identification of significant associations, each significant species pair was tested for differences in site-type categories (depending on aggregation or segregation as explained) using MANOVA. The test for environmental effects was based on 6 environmental principal components, while the test for spatial effects was based on 6 spatial eigenvectors, both retained from above. A significant ($\alpha = 0.05$) environmental effect is hypothesized to mean environmental filtering is the mechanism underlying the association. A significant ($\alpha = 0.05$) spatial effect is

hypothesized to mean dispersal limitation is the driving force. In the case that both spatial and environmental effects are significant it is likely that spatially autocorrelated habitat filtering is responsible. If neither is significant, then it is not possible to reject the hypothesis of a biotic interaction underlying the association. I considered marginally significant ($p <0.1$) effects as enough evidence of spatial or environmental signal so as to reduce associations that were attributed to biotic interactions (see Supplementary Material S.5-6 for full list of results).

R code is available in the Supplementary Material of this paper as R-scripts: "null-model_framework.R" and "probabilistic_framework.R". In addition, all data files including species incidence plus environment factors, shapefiles, and fluvial distance matrix are available in the supplementary material.

2.7 | Ethical Statement

The collection of animals complied with the University of Oklahoma Institutional Animal Care and Use Committee guidelines and policies (OU-IACUC approval No. R14-002). All specimen collecting was permitted by the Oklahoma Department of Wildlife Conservation (No. 5995 & No. 6166).

3 | Results

Twenty-four of the same native freshwater fish species were collected in both sampling periods and were present in at least 6 samples in at least one sampling period (=276 unique species pairs analyzed). No matter the year or framework, most species associations were not different from random (Table 1). The PRF resulted in more non-random species associations in both years

(Table 1). Results from NMF were skewed towards segregating associations, while those resulting from PRF were mostly aggregations (Table 1). For 1974 – 5 of 10 associations from the NMF were also recovered with PRF; for 2014, it was only 1 of 4 associations. The associations found exclusively by NMF in either year were all segregations. Associations from each year and framework combination are available in the Supplementary Material (S1-S4).

Fifty-four unique non-random species associations (19.6%) were found across years and frameworks (Supplementary Material: S5-S6). Of these associations, 22 (41%) were between species of the same family. Out of 24 species, only 2 were not involved in a single significant non-random association: slough darter *Etheostoma gracile* (Girard 1859) and orangespotted sunfish *Lepomis humilis* (Girard 1858). The number of associations per species ranged from 0 to 8, with a mean of 4.45 associations. The number of non-random species associations was not significantly associated with the number of occurrences per species (R^2=0.09, p=0.16) or rank occurrence (R^2=0.12, p=0.10) via linear regression (Figure 3).

Most non-random pairs were aggregated (42 pairs, 78%) compared to segregated (12 pairs, 22%). Most of these pairs were hypothesized as governed by non-biotic processes (28 pairs, 52%): habitat filtering (20 pairs, 37%), dispersal limitation (2 pairs, 0.4%), or both (6 pairs, 11%). The remaining 26 non-randomly associated pairs (48%) had no detectable difference in the appropriate site-type categories via MANOVA, therefore, the potential of biotic interaction was not refuted. For the 42 *aggregated* species pairs, 19 pairs (45%) were related to non-biotic processes: habitat filtering (13 pairs, 31%), dispersal limitation (2 pairs, 5%), or both (4 pairs, 10%). The remaining 23 (55%) *aggregated* pairs were potentially driven by biotic interactions.

For the 12 *segregated* species pairs, 9 (75%) were driven by non-biotic processes: habitat filtering (7 pairs; 58%), dispersal limitation (0 pair, 0%), or both space and habitat (2 pair, 17%). Only 3 of 12 *segregated* pairs (25%) were potentially driven by biotic interaction.

Among the 54 significantly non-random species pairs recovered across analyses, only 10 were recovered in multiple scenarios—in either both frameworks and/or across sampling periods (Table 2). Only 5 pairs of those pairs were found across both sampling periods. *Campostoma spp.* and orangebelly darter *Etheostoma radiosum* (Hubbs & Black 1941) were aggregated in both years with no discernable effect of dispersal limitation or habitat filtering. Red shiner *Cyprinella lutrensis* (Baird & Girard 1853) and bullhead minnow *Pimephales vigilax* (Baird & Girard 1853) were another aggregated pair found across years with an effect of habitat filtering found in both years. Bluegill sunfish *Lepomis macrochirus* Rafinesque 1819 and redear sunfish *Lepomis microlophus* (Günther 1859) were aggregated across times again without a signal of spatial or environmental effect. Redfin shiner *Lythrurus umbratilis* (Girard 1856) and bluntnose minnow *Pimephales notatus* (Rafinesque 1820) were yet another aggregated pair with an effect of habitat filtering found in 1974 but not in 2014. Finally, bigeye shiner *Notropis boops* Gilbert 1884 and golden shiner *Notemigonus crysoleucas* (Mitchill 1814) were the only segregated pair found consistently across time periods and this association is likely due to spatially autocorrelated habitat filtering.

4 | Discussion

4.1 | Species co-occurrence

The results point to many associations among fish species in the Muddy Boggy River drainage. The frameworks used herein were able to identify 54 unique, significantly non-random associations from incidence data that contained 276 pairwise comparisons. However, the proportion of neutral associations (not different from random) was high across both time periods and frameworks. This is typical for studies of this kind (Lopes *et al.*, 2015; Lyons *et al.*, 2016) but see Cordero & Jackson (2019). Some authors have argued large proportion of random associations points to the importance of stochasticity in structuring assemblages (Devercelli *et al.*, 2016). While stochasticity may play an important role, especially at smaller spatial scales (Zbinden, 2020), there is also evidence for deterministic processes, such as environmental filtering, structuring fish assemblages (Cordero & Jackson, 2019; Zbinden & Matthews, 2017). Some caution is warranted in rushing to conclusions regarding stochasticity (Rahel, Lyons, & Cochran, 1984) from analyses of this kind because of alternative explanations which are not explicitly tested. First, by considering only presence-absence data, information is lost and there is potential for associations to go undetected. If, for example, two species sometimes co-occur (weak signal), but when they do one of the species is invariably less abundant than when found without the other. That evidence might suggest a competitive interaction. However, this information is lost when considering only incidence. Second, pairwise comparisons have limited power to detect associations involving rare species (Lavender *et al.*, 2019; Veech, 2013), which are often a large portion of diversity in a region (Lavoie *et al.*, 2009; Magurran & Henderson, 2003), due to low statistical power. Third, sample sizes typical of most biodiversity surveys are likely not large enough to reliably capture all of the associations that may be present, with some estimates suggesting upwards of 500 or more samples may be necessary to do so (Blanchet *et al.*, 2020). So, while stochasticity is an important mechanism (Ellwood *et al.*, 2009; Vellend *et al.*,

2014), we must not neglect that methods that test co-occurrence are likely to result in many "random" associations simply due to their limitations independent of stochasticity. Importantly, we should not neglect that an "assemblage" is also only a part of the larger community in which interactions could be observed.

4.2 | What about Mechanisms?

The decision tree framework for disentangling likely mechanisms that underly associations is constructed so that biotic interaction is essentially the null hypothesis. Unfortunately, that means biotic interaction can be refuted, and alternatives supported, but in the absence of a signal from either spatial or environmental factors then the *de facto* explanation for a significant association is biotic interaction. Therefore, I emphasize that this inference should be viewed as hypothesis generation rather than hypothesis testing *per se*. Species associations that remain plausibly driven by biotic interaction should be judged by those conducting the analyses in order to determine whether the categorization makes ecological sense and warrants further investigation. It remains possible that unmeasured environmental factors could be involved in any of the associations identified. Additionally, the outcomes from the framework are not mutually exclusive (Blois *et al*., 2014), and the expected patterns from processes such as competition and predation may not always be straightforward (Blanchet *et al*., 2020; Cordero & Jackson, 2019).

Co-occurrence methods were largely developed with classic species interactions in mind (e.g., competition and predation). However, the results here demonstrated that most significant associations, gleaned from incidence data gathered across this study drainage, could be explained by non-biotic mechanisms, namely habitat filtering, but also dispersal limitation. It is intuitive

that species which differ in resource requirements (i.e., niche divergence) are less likely to be found in the same stream reaches, while those with similar requirements (i.e., niche convergence) would more often co-occur. Indeed spatio-environmental filtering (Cordero & Jackson, 2019; Peres-Neto, 2004; Zbinden & Matthews, 2017) is a well-supported hypothesis of metacommunity theory (Erős, 2017; Leibold *et al*., 2004). While habitat filtering is interesting in and of itself, if one is specifically seeking biotic interactions, then these frameworks can reveal which associations are likely due to alternative explanations instead.

No species were inferred to be segregated due to spatial factors alone (i.e., dispersal limitation); perhaps the relatively small spatial scale of this study is not large enough to truly segregate species by distance (Heino *et al*., 2015). Dispersal limitation was indicated, however, in the aggregation among *Campostoma spp.* and *Pimephales notatus* and among *E. radiosum* and freckled madtom *Noturus nocturnus* Jordan & Gilbert 1886. This suggests a pattern of nested, not entirely overlapping, distributions among species within the pair. Six interactions showed significant signal of both spatial and environmental factors in structuring the association and likely indicates the effect of spatially autocorrelated habitat filtering (Gotelli & Ulrich, 2012; Mykrä, Heino, & Muotka, 2007). In other words, the environmental factors associated with the species interactions were themselves spatially structured.

Only a small subset of the associations found here were segregations that remained plausibly driven by negative biotic interactions which would fit the prediction for pairs of competitors or predator/prey—although predator/prey relationships could result in aggregations as well (Cordero & Jackson, 2019), especially at the spatial scale analyzed here (Hoeinghaus,

Winemiller, & Birnbaum, 2007). These pairs were spotted bass *Micropterus punctulatus* (Rafinesque 1819) and *Notemigonus crysoleucas*, *E. radiosum* and white crappie *Pomoxis annularis* Rafinesque 1818, and *Lythrurus umbratilis* and *Notemigonus crysoleucas*. Both *M. punctulatus* and *Pomoxis annularis* are large centrarchid predators which could easily consume the other member of their respective pair. *Notemigonus crysoleucas* is a common forage fish consumed by many fish predators (Webber & Haines, 2003) and *E. radiosum* is a small benthic percid, although the only explicit test of predation on *E. radiosum* showed a lack of piscine predation (Scalet, 1974). *Lythrurus umbratilis* and *Notemigonus crysoleucas* are the only association that seems plausibly driven by competition. Another study in Oklahoma also found no co-occurrence among these two species at localities along a small creek (Stewart, Knight, & Cashner, 1992). Both species occur in similar habitats and have similar diets and feeding modes (Robison & Buchanan, 2020) and so competition seems plausible. It is of note, that of the 12 segregated associations recovered here, *Notemigonus crysoleucas* was a member of 6 of them. This species is highly tolerant of temperature and turbidity extremes making it a good competitor (Robison & Buchanan, 2020). Overall, I find very little evidence of competition among the fish species analyzed in this drainage, although it should be noted that only strongly competing species are likely to result in strong and significant segregated patterns. Furthermore, others have found competing species to sometimes aggregate rather than segregate, although this was at a much larger spatial scale (Cordero & Jackson, 2019).

A large portion of non-random aggregations (23 pairs) showed no discernable effect of spatial or environmental factors given the scale and environmental characteristics considered in this study. Therefore, the possibility of positive, or facilitative, interactions—but also associations due to

predator prey interactions—driving these associations remains but should be scrutinized. Seven of these 23 pairs included two members of the sunfish and black bass family Centrarchidae, within which individuals display niche flexibility based on size that often leads to what is referred to as "niche packing" or "species packing" (Desselle, Poirrier, Rogers, & Cashner, 1978; E. E. Werner & Hall, 1976; Werner, 1977). In effect, these species forage on similar resources, but diverge on the specific size of the resources targeted. Target prey size is normally correlated with predator size and mouth morphology, and this allows for stable coexistence of predators. However, species packing is not mutually exclusive from facilitation, and there is evidence that piscivorous predators using different strategies can benefit from one another (Eklöv & VanKooten, 2001; Schumann, Colvin, Miranda, & Jones-Farrand, 2020).

An additional 7 non-random pairs of those showing no discernable effect of spatial or environmental factors consisted of just one member of the Centrarchidae family plus a non-sunfish member (the other member being a minnow Leuciscidae in 5 of those cases). This may be explained, in part, by spawning nest associations (Johnson & Page, 1992; Peoples & Frimpong, 2016; Wallin, 1989). Members of the sunfish family make large nests by clearing depressions in stream-bottom substrate which are often used by other species (Johnston, 1994a; Stephen T. Ross, 2013) which benefit from the protective parental care of the nest-builder and not necessarily the physical environment of the nest (Johnston, 1994b). Examples exist of minnow species parasitizing centrarchid nests, sometimes even predating largely on the hosts eggs (Fletcher, 1993) or otherwise lowering host fitness through fungal infection of its brood (Shao, 1997). All species in this group of 7 pairs show strong overlap in spawning season (April – August in this region), and two of the species have been reported laying eggs in sunfish nests:

Cyprinella lutrensis and *Lythrurus umbratilis* (Robison & Buchanan, 2020 and references therein). Any of these pairs not engaged in nest association could possibly represent a predator prey interaction between the larger Centrarchid predator and smaller prey species (Hoeinghaus et al., 2007). Additional data would be required to parse these alternative possibilities, along with the potential effect of unmeasured environmental factors.

Campostoma spp. and *Etheostoma radiosum* is an interesting association recovered here because these species were significantly aggregated in 3 of 4 models across both sampling periods (indicating a reliable association), and yet could never be associated with dispersal limitation or habitat filtering. *Campostoma spp.* occurred in 16/33 samples (1974) and 22/33 samples (2014). *E. radiosum* occurred in 19/33 samples (1974) and 21/33 samples (2014). These two co-occurred together in 13 and 18 samples in 1974 and 2014, respectively. This result suggests a facilitative relationship between *Campostoma spp.* and *E. radiosum* may exist. *Campostoma spp.* are benthic, algivorous minnows that forage in shoals of up to many thousands of individuals (Matthews, Stewart, & Power, 1987), and can have important effects on other trophic levels within their communities such as producers and decomposers (Power *et al*., 1985; Veach, Troia, Jumpponen, & Dodds, 2018). Grazing fish like *Campostoma* can also promote the downstream drift of macroinvertebrates by disturbing substrate and causing these organisms to enter the water column. Matthews *et al*. (1987) discussed the potential for *Campostoma* to have positive interactions on other fish by improving feeding opportunities for macroinvertebrate predators through substrate disturbance, and provided anecdotal evidence of an association between duskystripe shiner *Luxilus pilsbryi* (Fowler 1904) and *Campostoma* that often occurred in the same pools together as observed during snorkeling surveys. A similar association has been

reported between *Campostoma anomalum* and hornyhead chub *Necomis biguttatus* (Kirtland 1840) (Gorman, 1988). Various suckers Catostomidae are known feeding facilitators that have a similar feeding mode that involves disturbing substrate (Pflieger, 1975; Reighard, 1920) as well as various other species of freshwater fish (Arnhold, Penha, Peoples, & Mateus, 2019; Stephen T. Ross, 2013). Therefore, it is possible that large shoals of foraging *Campostoma* disturb substrate and force more macroinvertebrates into the water column which then drift downstream into riffles occupied by darters, like *E. radiosum,* that benefit from higher abundance of food. This is an intriguing hypothesis that deserves future consideration, as the importance of large populations of algivorous grazers is appreciated by fish ecologists (Matthews et al., 1987), but perhaps not fully understood.

4.3 | Inconsistency through time

Similarity in the species associations recovered across sampling campaigns was low (only 5 consistent associations). Species interactions may be expected to change over time due to both species turnover and "rewiring" of interactions (Brimacombe et al., 2020). Species may associate only transiently (Bart, 1989; Grossman & Ratajczak, 1998), seasonally (Gido & Matthews, 2000; Matthews & Hill, 1980), or over long periods of time (Matthews & Matthews, 2002; Matthews & Marsh-Matthews, 2016; Ross, Matthews, & Echelle, 1985). Species associations may shift over time due to many non-mutually exclusive factors including environmental change, population movement and demographic shifts, and local extinctions, all of which are all scale-dependent processes (Blanchet *et al*., 2020; Chafin, Douglas, Martin, & Douglas, 2019; Harmon, Moran, & Ives, 2009).

A previous comparison of fish assemblage structure in the Muddy Boggy River drainage over the same period measured high temporal turnover at the local spatial scale (i.e. sampling locality) (Zbinden, 2020). This and the similarity in season among the two sampling campaigns might suggest the importance of turnover as opposed to interaction rewiring (Brimacombe et al., 2020) for explaining the low similarity in species association through time. As Lavender *et al.* (2019) showed, changes in species incidence can affect the ability of these tests to determine non-random interactions. These authors cautioned against comparing tests among sampling times or experimental treatments because of this reason. I take this point; however, I would argue not against comparison, but for more interest in similarities rather than differences. For the study presented here, many dissimilarities were found between sampling times and some of these may be due to changes in local incidences over time, among a host of other factors. However, non-random associations that persist through time (as discussed above) provide stronger and more reliable evidence for two species associating.

4.4. | Complimentary frameworks

The two co-occurrence frameworks used here appear to complement each other well, because each provided unique associations not identified by the other. The Probabilistic Framework (PRF) and Null Model Framework (NMF) both aim to identify non-randomly associated species pairs, but they differ in statistical approach. Null model approaches with Bayes corrections had one of the lowest Type I error rates among co-occurrence methods, but consistently failed to detect known associations from simulated data in a study of co-occurrence methods (Lavender *et al.,* 2019). The Probabilistic model (Veech, 2013) used here actually performed best overall with regard to both error rates in that same comparison (Lavender *et al*., 2019). Therefore, I expected

a priori that the NMF would result in fewer significant association compared to the PRF, and indeed this was the case. Lavender *et al.* (2019) also found that negative (segregated associations) had lower Type I and higher Type II error rates. Generally, null methods seem to be more robust tests of negative associations, while probabilistic methods are more robust tests of positive associations. The findings here are also consistent with this result as there were many more aggregated than segregated pairs.

Despite low Type I error, methods used here must be interpreted with caution. For example, this study analyzed 276 unique species pairs, and given an error rate of ~ 0.01, it is possible that ~ 3 or less of the significantly non-random pairs were false positives. Future studies could also reduce error rate by only analyzing species hypothesized *a priori* (e.g. *Campostoma spp.* and *Etheostoma spp.*) to have an association, or by reducing the set to include only certain species of interest (e.g. species of conservation concern, species within families, or known predator/prey species).

4.5 | Conclusion

Different species can associate or interact in many ways, and methods exist for inferring associations and underlying mechanisms from incidence data (e.g., co-occurrence frameworks). However, these methods have received much criticism despite their recent resurgence in the literature. Co-occurrence frameworks for identifying non-randomly associated species pairs (e.g. aggregated or segregated pairs) should be used as heuristic tools for sharpening hypotheses concerning fish ecology (Silknetter et al., 2020). The data presented here demonstrated that at least some species associations can be found in even a small number of samples. Roughly half of

these associations were driven by non-biotic processes (e.g., habitat filtering). And most associations did not show a consistent signal through time.

All the significant associations from this study can be explored further. Out of 276 possible associations, there were 54 that showed a non-random pattern. These can further be prioritized based on the number of models or times they were significant. These frameworks can provide insight into the broader mechanisms of species assembly and point to potentially interesting species interactions (out of many possible interactions). Rather than a weakness, the heuristic nature is the strength of such methods, and can help guide biologists toward better questions by employing relatively cheap diversity survey data, that are often already in hand, so as to reduce complex interaction networks down to their non-stochastic parts for future investigation. This can save fisheries biologists time and resources and help guide us toward species associations that warrant further investigation. In turn, this can lead to a fuller understanding of how fish assemblages are structured.

Data Availability Statement

All data and R code available in the supplementary material.

Significance Statement

Species interactions and mechanisms are of interest to fisheries biologists if we hope to successfully conserve populations. While many species do not interact, those that do may be identified with pairwise analysis of species association (i.e. finding the "needle in the haystack"). Further testing can also help sharpen hypotheses regarding the mechanism(s) that drive these interactions. Although these methods have faced scrutiny, that is largely because of how those who employ them approach their use. Rather than "definitive", these tests should be thought of as tools for sharpening hypotheses and focusing on research questions that matter. Identifying target interactions and potential explanations can focus further empirical study and save time and resources spent on identifying the important biological connections in communities.

References

Aragón, P., & Sánchez-Fernández, D. (2013). Can we disentangle predator-prey interactions from species distributions at a macro-scale? A case study with a raptor species. *Oikos*, *122*(1), 64–72. https://doi.org/10.1111/j.1600-0706.2012.20348.x

Arnhold, T. R., Penha, J., Peoples, B. K., & Mateus, L. A. F. (2019). Positive co-occurrence between feeding-associative savannah fishes depends on species and habitat. *Freshwater Biology*, *64*(5), 1029–1039. https://doi.org/10.1111/fwb.13283

Barner, A. K., Coblentz, K. E., Hacker, S. D., & Menge, B. A. (2018). Fundamental contradictions among observational and experimental estimates of non-trophic species

interactions. *Ecology*, *99*(3), 557–566. https://doi.org/10.1002/ecy.2133

Bart, H. L. (1989). Fish habitat association in an Ozark stream. *Environmental Biology of Fishes*, *24*(3), 173–186. https://doi.org/10.1007/BF00001222

Bell, G. (2005). The co-distribution of species in relation to the neutral theory of community ecology. *Ecology*, *86*(7), 1757–1770. https://doi.org/10.1890/04-1028

Benton, M. J. (2009, February 6). The Red Queen and the Court Jester: Species diversity and the role of biotic and abiotic factors through time. *Science*. American Association for the Advancement of Science. https://doi.org/10.1126/science.1157719

Blanchet, F. G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*. https://doi.org/10.1111/ele.13525

Blois, J. L., Gotelli, N. J., Behrensmeyer, A. K., Faith, J. T., Lyons, S. K., Williams, J. W., … Wing, S. (2014). A framework for evaluating the influence of climate, dispersal limitation, and biotic interactions using fossil pollen associations across the late Quaternary. *Ecography*, *37*(11), 1095–1108. https://doi.org/10.1111/ecog.00779

Bonin, M. C., Boström, L., Boström-Einarsson, B., Munday, P. L., & Jones, G. P. (2015). The Prevalence and Importance of Competition Among Coral Reef Fishes. *Annu. Rev. Ecol. Evol. Syst*, *46*, 169–190. https://doi.org/10.1146/annurev-ecolsys-112414-054413

Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, *153*(1–2), 51–68. https://doi.org/10.1016/S0304-3800(01)00501-4

Brimacombe, C., Bodner, K., & Fortin, M. (2020). Inferred seasonal interaction rewiring of a freshwater stream fish network. *Ecography*, *43*, 1–12. https://doi.org/10.1111/ecog.05452

Calatayud, J., Andivia, E., Escudero, A., Melián, C. J., Bernardo-Madrid, R., Stoffel, M., …

Madrigal-González, J. (2020). Positive associations among rare species and their persistence in ecological assemblages. *Nature Ecology and Evolution*, *4*(1), 40–45. https://doi.org/10.1038/s41559-019-1053-5

Case, T. J., & Taper, M. L. (2000). Interspecific competition, environmental gradients, gene flow, and the coevolution of species' borders. *American Naturalist*, *155*(5), 583–605. https://doi.org/10.1086/303351

Cashner, R. C., Matthews, W. J., Marsh-Matthews, E., Unmack, P. J., & Cashner, F. M. (2010). Recognition and Redescription of a Distinctive Stoneroller from the Southern Interior Highlands. *Copeia*, *2010*(2), 300–311. https://doi.org/10.1643/CI-08-051

Cazelles, K., Araújo, M. B., Mouquet, N., & Gravel, D. (2016). A theory for species co-occurrence in interaction networks. *Theoretical Ecology*, *9*(1), 39–48. https://doi.org/10.1007/s12080-015-0281-9

Chafin, T. K., Douglas, M. R., Martin, B. T., & Douglas, M. E. (2019). Hybridization drives genetic erosion in sympatric desert fishes of western North America. *Heredity*, *123*(6), 759–773. https://doi.org/10.1038/s41437-019-0259-2

Connor, E. F., & Simberloff, D. (1979). The Assembly of Species Communities: Chance or Competition? *Ecology*, *60*(6), 1132–1140. https://doi.org/10.2307/1936961

Cordero, R. D., & Jackson, D. A. (2019). Species-pair associations, null models, and tests of mechanisms structuring ecological communities. *Ecosphere*, *10*(7). https://doi.org/10.1002/ecs2.2797

D'Amen, M., Mod, H. K., Gotelli, N. J., & Guisan, A. (2018). Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species co-occurrence. *Ecography*, *41*(8), 1233–1244. https://doi.org/10.1111/ecog.03148

Desselle, W. J., Poirrier, M. A., Rogers, J. S., & Cashner, R. C. (1978). A Discriminant

Functions Analysis of Sunfish (Lepomis) Food Habits and Feeding Niche Segregation in

the. *Transactions of the American Fisheries Society*, *107*(5), 713–719.

https://doi.org/10.1577/1548-8659(1978)107<713:ADFAOS>2.0.CO;2

Devercelli, M., Scarabotti, P., Mayora, G., Schneider, B., & Giri, F. (2016). Unravelling the role

of determinism and stochasticity in structuring the phytoplanktonic metacommunity of the

Paraná River floodplain. *Hydrobiologia*, *764*(1), 139–156. https://doi.org/10.1007/s10750-

015-2363-5

Diamond, J. M. (1975). Assembly of species communities. In M. L. Cody & J. M. Diamond

(Eds.), *Ecology and Evolution of communities* (pp. 342–344). Harvard University Press.

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., … Wagner, H. (2020).

Package "adespatial": Multivariate Multiscale Spatial Analysis. Retrieved from

https://cran.r-project.org/package=adespatial

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive

framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological

Modelling*, *196*(3–4), 483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015

Efron, B. (2005). Bayesians, frequentists, and scientists. *Journal of the American Statistical

Association*, *100*(469), 1–5. https://doi.org/10.1198/016214505000000033

Eklöv, P., & VanKooten, T. (2001). Facilitation among piscivorous predators: Effects of prey

habitat use. *Ecology*, *82*(9), 2486–2494. https://doi.org/10.1890/0012-

9658(2001)082[2486:FAPPEO]2.0.CO;2

Erős, T. (2017). Scaling fish metacommunities in stream networks: synthesis and future research

avenues. *Community Ecology*, *18*(1), 72–86. https://doi.org/10.1556/168.2017.18.1.9

Farnon Ellwood, M. D., Manica, A., Foster, W. A., Ellwood, M. D. F., Manica, A., & Foster, W. A. (2009). Stochastic and deterministic processes jointly structure tropical arthropod communities. *Ecology Letters*, *12*(4), 277–284. https://doi.org/10.1111/j.1461-0248.2009.01284.x

Fauth, J., Bernardo, J., & Camara, M. (1996). Simplifying the Jargon of Community Ecology : A Conceptual Approach. *The American Naturalist*, *147*(2), 282–286. https://doi.org/10.1086/285850

Fletcher, D. E. (1993). Nest Association of Dusky Shiners (Notropis cummingsae) and Redbreast Sunfish (Lepomis auritus), a Potentially Parasitic Relationship. *Copeia*, *1993*(1), 159. https://doi.org/10.2307/1446306

Forbes, S. A. (1907). On the local distribution of certain Illinois fishes: an essay in statistical ecology. *Bulletin of the Illinois State Laboratory of Natural History*, *7*(8).

Freilich, M. A., Wieters, E., Broitman, B. R., Marquet, P. A., & Navarrete, S. A. (2018). Species co-occurrence networks: Can they reveal trophic and non-trophic interactions in ecological communities? *Ecology*, *99*(3), 690–699. https://doi.org/10.1002/ecy.2142

Geheber, A. D., & Frenette, B. D. (2016). Separation in habitat use and phylogenetic influence on habitat use among fishes in diverse temperate stream communities. *Hydrobiologia*, *767*(1), 235–248. https://doi.org/10.1007/s10750-015-2504-x

Gido, K. B., & Matthews, W. J. (2000). Dynamics of the offshore fish assemblage in a southwestern reservoir (Lake Texoma, Oklahoma-Texas). *Copeia*, *2000*(4), 917–930. https://doi.org/10.1643/0045-8511(2000)000[0917:DOTOFA]2.0.CO;2

Gilman, S. E., Urban, M. C., Tewksbury, J., Gilchrist, G. W., & Holt, R. D. (2010). A framework for community interactions under climate change. *Trends in Ecology and Evolution*, *25*(6),

325–331. https://doi.org/10.1016/j.tree.2010.03.002

Gilpin, M. E., & Diamond, J. M. (1982). Factors contributing to non-randomness in species Co-occurrences on Islands. *Oecologia*, *52*(1), 75–84. https://doi.org/10.1007/BF00349014

Gorman, O. T. (1988). The Dynamics of Habitat Use in a Guild of Ozark Minnows. *Ecological Monographs*, *58*(1), 1–18. https://doi.org/10.2307/1942631

Gotelli, N. J. (2000). Null Model Analysis of Species Co-Occurrence Patterns. *Ecology*, *81*(9), 2606–2621. https://doi.org/10.2307/177478

Gotelli, N. J., & McCabe, D. J. (2002). Species co-occurrence: A meta-analysis of J. M. Diamond's assembly rules model. *Ecology*, *83*(8), 2091–2096. https://doi.org/10.1890/0012-9658(2002)083[2091:SCOAMA]2.0.CO;2

Gotelli, N. J., & Ulrich, W. (2010). The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia*, *162*(2), 463–477. https://doi.org/10.1007/s00442-009-1474-y

Gotelli, N. J., & Ulrich, W. (2012, February). Statistical challenges in null model analysis. *Oikos*. https://doi.org/10.1111/j.1600-0706.2011.20301.x

Griffith, D. M., Veech, J. A., & Marsh, C. J. (2016). cooccur : Probabilistic Species Co-Occurrence Analysis in R. *Journal of Statistical Software*, *69*(2), 1–17. https://doi.org/10.18637/jss.v069.c02

Grossman, G. D., & Ratajczak, R. E. (1998). Long-term patterns of microhabitat use by fish in a southern Appalachian stream from 1983 to 1992: effects of hydrologic period, season and fish length. *Ecology of Freshwater Fish*, *7*(3), 108–131. https://doi.org/10.1111/j.1600-0633.1998.tb00178.x

Harmon, J. P., Moran, N. A., & Ives, A. R. (2009). Species Response to Environmental Change:

Impacts of Food Web Interactions and Evolution. *Science*, *323*(5919), 1347–1350. https://doi.org/10.1126/science.1167396

Heino, J., & Grönroos, M. (2013). Does environmental heterogeneity affect species co-occurrence in ecological guilds across stream macroinvertebrate metacommunities? *Ecography*, *36*(8), 926–936. https://doi.org/10.1111/j.1600-0587.2012.00057.x

Heino, J., Melo, A. S., Siqueira, T., Soininen, J., Valanko, S., & Bini, L. M. (2015). Metacommunity organisation, spatial extent and dispersal in aquatic systems: Patterns, processes and prospects. *Freshwater Biology*, *60*(5), 845–869. https://doi.org/10.1111/fwb.12533

Hijmans, R. J., & van Etten, J. (2015). raster: Geographic analysis and modeling with raster data. *R Package Version 2.5-2*.

Hoeinghaus, D. J., Winemiller, K. O., & Birnbaum, J. S. (2007). Local and regional determinants of stream fish assemblage structure: Inferences based on taxonomic vs. functional groups. *Journal of Biogeography*, *34*(2), 324–338. https://doi.org/10.1111/j.1365-2699.2006.01587.x

Holt, R. D., & Bonsall, M. B. (2017). Apparent Competition. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 447–471. https://doi.org/10.1146/annurev-ecolsys-110316-022628

Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. *Monographs in Population Biology*. Princeton University Press. https://doi.org/10.1016/S0006-3207(02)00228-8

Jackson, D. A., Peres-Neto, P. R., & Olden, J. D. (2001). What controls who is where in freshwater fish communities – the roles of biotic, abiotic, and spatial factors. *Canadian*

Journal of Fisheries and Aquatic Sciences, *58*(1), 157–170. https://doi.org/10.1139/cjfas-58-1-157

Johnson, C. E., & Page, L. M. (1992). The evolution of complex reproductive strategies in North American minnows. In R. L. Mayden (Ed.), *Systematics, historical ecology, and North American freshwater fishes* (pp. 600–621). Stanford, CA: Standford University Press.

Johnston, C. E. (1994a). Nest association in fishes: evidence for mutualism. *Behavioral Ecology and Sociobiology*, *35*(6), 379–383. https://doi.org/10.1007/BF00165839

Johnston, C. E. (1994b). The benefit to some minnows of spawning in the nests of other species. *Environmental Biology of Fishes*, *40*(2), 213–218. https://doi.org/10.1007/BF00002547

Lavender, T. M., Schamp, B. S., Arnott, S. E., & Rusak, J. A. (2019). A comparative evaluation of five common pairwise tests of species association. *Ecology*, *100*(4), 1–8. https://doi.org/10.1002/ecy.2640

Lavoie, I., Dillon, P. J., & Campeau, S. (2009). The effect of excluding diatom taxa and reducing taxonomic resolution on multivariate analyses and stream bioassessment. *Ecological Indicators*, *9*(2), 213–225. https://doi.org/10.1016/j.ecolind.2008.04.003

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., … Gonzalez, A. (2004, July). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*. https://doi.org/10.1111/j.1461-0248.2004.00608.x

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., … Thieme, M. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Scientific Data*, *6*(1), 283. https://doi.org/10.1038/s41597-019-0300-6

Lopes, G. N., Souza-Filho, M. F., Gotelli, N. J., Lemos, L. J. U., Godoy, W. A. C., & Zucchi, R. A. (2015). Temporal overlap and co-occurrence in a guild of sub-tropical tephritid fruit

flies. *PLoS ONE*, *10*(7). https://doi.org/10.1371/journal.pone.0132124

Lyons, S. K., Amatangelo, K. L., Behrensmeyer, A. K., Bercovici, A., Blois, J. L., Davis, M., …
Gotelli, N. J. (2016). Holocene shifts in the assembly of plant and animal communities
implicate human impacts. *Nature*, *529*(7584), 80–83. https://doi.org/10.1038/nature16447

MacKenzie, D. I., Bailey, L. L., & Nichols, J. D. (2004). Investigating species co-occurrence
patterns when species are detected imperfectly. *Journal of Animal Ecology*, *73*(3), 546–555.
https://doi.org/10.1111/j.0021-8790.2004.00828.x

Magurran, A. E., & Henderson, P. A. (2003). Explaining the excess of rare species in natural
species abundance distributions. *Nature*, *422*(6933), 714–716.
https://doi.org/10.1038/nature01547

Matthews, E., & Matthews, W. J. (2002). Temporal Stability of Minnow Species Co-Occurrence
in Streams of the Central United States. *Transactions of the Kansas Academy of Science*,
105(3 & 4), 162–177. https://doi.org/10.1660/0022-
8443(2002)105[0162:tsomsc]2.0.co;2

Matthews, W.J. (1998). *Patterns in Freshwater Fish Ecology*. New York: Chapman and Hall.

Matthews, W.J., & Marsh-Matthews, E. (2016). Dynamics of an upland stream fish community
over 40 years: Trajectories and support for the loose equilibrium concept. *Ecology*, *97*(3),
706–719. https://doi.org/10.1890/14-2179

Matthews, W.J., & Marsh-Matthews, E. (2017). *Stream fish community dynamics*. Baltimore,
MD: Johns Hopkins University Press.

Matthews, W.J., Stewart, A. J., & Power, M. E. (1987). Grazing Fishes as Components of North
American Stream Ecosystems: Effects of Campostoma anomalum. In W.J. Matthews & D.
C. Heins (Eds.), *Community and evolutionary ecology of North American stream fishes* (pp.

128–135). Norman, OK: Univeristy of Oklahoma Press.

Matthews, William J., & Hill, L. G. (1980). Habitat Partitioning in the Fish Community of a Southwestern River. *The Southwestern Naturalist*, *25*(1), 51. https://doi.org/10.2307/3671211

McMeans, B. C., McCann, K. S., Guzzo, M. M., Bartley, T. J., Bieg, C., Blanchfield, P. J., … Shuter, B. J. (2020). Winter in water: differential responses and the maintenance of biodiversity. *Ecology Letters*, *23*(6), 922–938. https://doi.org/10.1111/ele.13504

Morales-Castilla, I., Matias, M. G., Gravel, D., & Araújo, M. B. (2015, June 1). Inferring biotic interactions from proxies. *Trends in Ecology and Evolution*. Elsevier Ltd. https://doi.org/10.1016/j.tree.2015.03.014

Mykrä, H., Heino, J., & Muotka, T. (2007). Scale-related patterns in the spatial and environmental components of stream macroinvertebrate assemblage variation. *Global Ecology and Biogeography*, *16*(2), 149–159. https://doi.org/10.1111/j.1466-8238.2006.00272.x

Pebesma, E. (2019). Package 'sf' - Simple Features for R. *CRAN Repository*.

Peoples, B. K., & Frimpong, E. A. (2016). Biotic interactions and habitat drive positive co-occurrence between facilitating and beneficiary stream fishes. *Journal of Biogeography*, *43*(5), 923–931. https://doi.org/10.1111/jbi.12699

Peres-Neto, P. R. (2004). Patterns in the co-occurrence of fish species in streams: The role of site suitability, morphology and phylogeny versus species interactions. *Oecologia*, *140*(2), 352–360. https://doi.org/10.1007/s00442-004-1578-3

Perry, G. L. W., Miller, B. P., Lamont, B. B., & Enright, N. J. (2020). Matrix-level co-occurrence metrics are sensitive to sampling grain: a case study in species-rich shrublands.

Plant Ecology, *221*(11), 1083–1090. https://doi.org/10.1007/s11258-020-01063-8

Peterson, A. T., Soberón, J., Ramsey, J., & Osorio-Olvera, L. (2020). Co-occurrence Networks do not Support Identification of Biotic Interactions. *Biodiversity Informatics*, *15*(1), 1–10. https://doi.org/10.17161/bi.v15i1.9798

Pflieger, W. L. (1975). *The Fishes of Missouri*. Jefferson City: Missouri Department of Conservation.

Pielou, D. P., & Pielou, E. C. (1968). Association among species of infrequent occurrence: The insect and spider fauna of Polyporus betulinus (Bulliard) Fries. *Journal of Theoretical Biology*, *21*(2), 202–216. https://doi.org/10.1016/0022-5193(68)90070-2

Pigg, J. (1977). A survey of the fishes of the Muddy Boggy River in south central Oklahoma. *Proceedings of the Oklahoma Academy of Science*, *57*, 68–82.

Power, M. E., Matthews, W. J., & Stewart, A. J. (1985). Grazing minnows, piscivorous bass, and stream algae: dynamics of a strong interaction. *Ecology*, *66*(5), 1448–1456. https://doi.org/10.2307/1938007

Rahel, F. J., Lyons, J. D., & Cochran, P. A. (1984). Stochastic or deterministic regulation of assemblage structure? It may depend on how the assemblages is defined. *The American Naturalist*, *124*(4), 583–589. https://doi.org/10.1086/284297

RCoreTeam. (2019). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*.

Reighard, J. (1920). The breeding behavior of the suckrs and minnows. *The Biological Bulletin1*, *XXXVIII*, 1–32.

Robison, H. W., & Buchanan, T. M. (2020). *Fishes of Arkansas* (2nd ed.). Fayetteville, AR: University of Arkansas Press.

Ross, S T, Matthews, W. J., & Echelle, A. A. (1985). Persistence of stream fish assemblages: Effects of environmental change. *The American Naturalist*, *126*(1), 24–40. https://doi.org/10.1086/284393

Ross, Stephen T. (2013). *Ecology of North American Freshwater Fishes*. *Book*. https://doi.org/10.1111/jfb.12536

Scalet, C. G. (1974). Lack of Piscine Predation on the Orangebelly Darter, Etheostoma radisoum cyanorum. *The American Midland Naturalist*, *92*(2), 510–512.

Schamp, B. S., Arnott, S. E., & Joslin, K. L. (2015). Dispersal strength influences zooplankton co-occurrence patterns in experimental mesocosms. *Ecology*, *96*(4), 1074–1083. https://doi.org/10.1890/14-1128.1

Schemske, D. W., Mittelbach, G. G., Cornell, H. V., Sobel, J. M., & Roy, K. (2009). Is there a latitudinal gradient in the importance of biotic interactions? *Annual Review of Ecology, Evolution, and Systematics*, *40*, 245–269. https://doi.org/10.1146/annurev.ecolsys.39.110707.173430

Schluter, D. (1994). Experimental evidence that competition promotes divergence in adaptive radiation. *Science*, *266*(5186), 798–801. https://doi.org/10.1126/science.266.5186.798

Schumann, D. A., Colvin, M. E., Miranda, L. E., & Jones-Farrand, D. T. (2020). Occurrence and Co-Occurrence Patterns of Gar in River–Floodplain Habitats: Methods to Leverage Species Coexistence to Benefit Distributional Models. *North American Journal of Fisheries Management*, *40*(3), 622–637. https://doi.org/10.1002/nafm.10402

Schwilk, D. W., & Ackerly, D. D. (2005). Limiting similarity and functional diversity along environmental gradients. *Ecology Letters*, *8*(3), 272–281. https://doi.org/10.1111/j.1461-0248.2004.00720.x

Sfenthourakis, S., Tzanatos, E., & Giokas, S. (2006). Species co-occurrence: The case of congeneric species and a causal approach to patterns of species association. *Global Ecology and Biogeography*, *15*(1), 39–49. https://doi.org/10.1111/j.1466-822X.2005.00192.x

Shao, B. (1997). Effects of golden shiner (Notemigonus crysoleucas) nest association on host pumpkinseeds (Lepomis gibbosus): Evidence for a non-parasitic relationship. *Behavioral Ecology and Sociobiology*, *41*(6), 399–406. https://doi.org/10.1007/s002650050401

Silknetter, S., Creed, R. P., Brown, B. L., Frimpong, E. A., Skelton, J., & Peoples, B. K. (2020). Positive biotic interactions in freshwaters: A review and research directive. *Freshwater Biology*, *65*(4), 811–832. https://doi.org/10.1111/fwb.13476

Stachowicz, J. J. (2001). Mutualism, facilitation, and the structure of ecological communities. *BioScience*, *51*(3), 235–246. https://doi.org/10.1641/0006-3568(2001)051[0235:MFATSO]2.0.CO;2

Stewart, B. G., Knight, J. G., & Cashner, R. C. (1992). Longitudinal Distribution and Assemblages of Fishes of Byrd's Mill Creek, a Southern Oklahoma Arbuckle Mountain Stream. *The Southwestern Naturalist*, *37*(2), 138. https://doi.org/10.2307/3671662

Stone, L., & Roberts, A. (1990). Oecologia The checkerboard score and species distributions. *Oecologia*, *85*, 74–79.

Thurman, L. L., Barner, A. K., Garcia, T. S., & Chestnut, T. (2019). Testing the link between species interactions and species co-occurrence in a trophic network. *Ecography*, *42*(10), 1658–1670. https://doi.org/10.1111/ecog.04360

Ulrich, W. (2004). Species co-occurrences and neutral models: Reassessing J. M. Diamond's assembly rules. *Oikos*, *107*(3), 603–609. https://doi.org/10.1111/j.0030-1299.2004.12981.x

Veach, A. M., Troia, M. J., Jumpponen, A., & Dodds, W. K. (2018). Top–down effects of a

Accepted Article

grazing, omnivorous minnow (*Campostoma anomalum*) on stream microbial communities. *Freshwater Science*, *37*(1), 121–133. https://doi.org/10.1086/696292

Veech, J. A. (2013). A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, *22*(2), 252–260. https://doi.org/10.1111/j.1466-8238.2012.00789.x

Veech, J. A. (2014). The pairwise approach to analysing species co-occurrence. *Journal of Biogeography*, *41*(6), 1029–1035. https://doi.org/10.1111/jbi.12318

Vellend, M., Srivastava, D. S., Anderson, K. M., Brown, C. D., Jankowski, J. E., Kleynhans, E. J., … Xue, X. (2014). Assessing the relative importance of neutral stochasticity in ecological communities. *Oikos*, *123*(12), 1420–1430. https://doi.org/10.1111/oik.01493

Wallin, J. E. (1989). Bluehead Chub (Nocomis leptocephalus) Nests Used by Yellowfin Shiners (Notropis lutipinnis). *Copeia*, *1989*(4), 1077. https://doi.org/10.2307/1446006

Webber, H. M., & Haines, T. A. (2003). Mercury effects on predator avoidance behavior of a forage fish, golden shiner (*Notemigonus crysoleucas*). *Environmental Toxicology and Chemistry*, *22*(7), 1556–1561. https://doi.org/10.1002/etc.5620220718

Werner, E. E., & Hall, D. J. (1976). Niche Shifts in Sunfishes: Experimental Evidence and Significance. *Science*, *191*(4225), 404–406.

Werner, Earl E. (1977). Species Packing and Niche Complementarity in Three Sunfishes. *The American Naturalist*, *111*(979), 553–578. https://doi.org/10.1086/283184

Wiens, J. A. (1977). On Competition and Variable Environments. *American Scientist*, *65*, 590–597.

Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., … Svenning, J. C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*,

88(1), 15–30. https://doi.org/10.1111/j.1469-185X.2012.00235.x

Zbinden, Z. D. (2020). Temporal dynamics of stream fish assemblages and the role of spatial

scale in quantifying change. *Ecology and Evolution*, *10*(2), 952–961.

https://doi.org/10.1002/ece3.5954

Zbinden, Z. D., & Matthews, W. J. (2017). Beta diversity of stream fish assemblages:

partitioning variation between spatial and environmental factors. *Freshwater Biology*, *62*(8),

1460–1471. https://doi.org/10.1111/fwb.12960

Zurell, D., Pollock, L. J., & Thuiller, W. (2018). Do joint species distribution models reliably

detect interspecific interactions from co-occurrence data in homogenous environments?

Ecography, *41*(11), 1812–1819. https://doi.org/10.1111/ecog.03315

Significance Statement

Species interactions and mechanisms are of interest to fisheries biologists if we hope to

successfully conserve populations. While many species do not interact, those that do may be

identified with pairwise analysis of species association (i.e. finding the "needle in the haystack").

Further testing can also help sharpen hypotheses regarding the mechanism(s) that drive these

interactions. Although these methods have faced scrutiny, that is largely because of how those

who employ them approach their use. Rather than "definitive", these tests should be thought of

as tools for sharpening hypotheses and focusing on research questions that matter. Identifying

target interactions and potential explanations can focus further empirical study and save time and

resources spent on identifying the important biological connections in communities.

Table 1. Summary of outcomes for species association modelling for two years (1974 & 2014) and two frameworks (Null Model Framework and Probabilistic Framework). For both years and analyses, the majority of the 276 unique species combinations were not significantly different from random expectation (Total Random). Non-random pairs are broken down into Total Aggregated and Total Segregated.

Year	Framework	Total Random	Total Non-random	Total Aggr.	Total Segr.
1974	NMF	266 (94%)	10 (4%)	3	7
	PRF	241 (87%	35 (13%)	31	4
2014	NMF	272 (99%)	4 (1%)	1	3
	PRF	260 (94%)	16 (6%)	15	1

Table 2. Significant species association that were recovered in at least 2 frameworks or sampling periods. (For full table of all 54 species associations see Supplementary Tables). Column headers indicate which two species are part of the interaction (Species 1 & Species 2), the hypothesized mechanism driving the association (Hyp. Mech), the nature of the association, either aggregated or segregated (Null Mod.), the number of models that returned this specific association as significant, 4 total models (No. Models), and the remining columns represent each model: null model 1974 (1974 NMF), probabilistic model 1974 (1974 PRF), null model 2014 (2014 NMF), and the probabilistic model 2014 (2014 PRF).

No.	Species 1	Species 2	Hyp. Mech	Null Mod.	No. Models	1974 NMF	1974 PRF	2014 NMF	2014 PRF
1	*Campostoma spp.*	*Etheostoma radiosum*	Biotic interaction	AG GR	3		X	X	X
2	*Cyprinella lutrensis*	*Cyprinella venusta*	Habitat filtering	SEG R	2	X	X		
3	*Cyprinella lutrensis*	*Pimephales vigilax*	Habitat filtering	AG GR	2		X		X
4	*Fundulus spp.*	*Micropterus punctulatus*	Spatial/Habitat filtering	AG GR	2	X	X		
5	*Lepomis macrochirus*	*Lepomis microlophus*	Biotic interaction	AG GR	2		X		X
6	*Lythrurus umbratilis*	*Pimephales notatus*	Habitat filtering	AG GR	2		X		X
7	*Micropterus punctulatus*	*Notemigonus crysoleucas*	Biotic interaction	SEG R	2	X	X		
8	*Notropis boops*	*Notemigonus*	Spatial/Habitat	SEG	2		X	X	

		crysoleucas	filtering	R			
9	*Noturus nocturnus*	*Percina sciera*	Biotic interaction	AG GR	2	X	X
10	*Percina sciera*	*Pimephales vigilax*	Habitat filtering	AG GR	2	X	X

Figure 1. Map of the Muddy Boggy River in southeastern, OK, USA. Sites where fish specimens were collected are marked with a hollow black circle and the sampling ID is noted next to each.

Figure 2. Logic tree for deducing inferences driving significant species associations from incidence data. Null Model and Probabilistic Frameworks each use the same logic tree to make inferences (illustrated here). The frameworks differ in how significant associations are determined. The Null Model Framework relies on randomly shuffling the data matrix to generate random expectations, and the Probabilistic Framework uses probabilistic combinatorics to quantify the expected probabilities of species aggregation and segregation based on the data set.

Figure 3. This scatterplot shows the relationship between the number of non-random associations and rank occurrence (based on number of occurrences across 66 samples) for each of the 24 species analyzed here.s
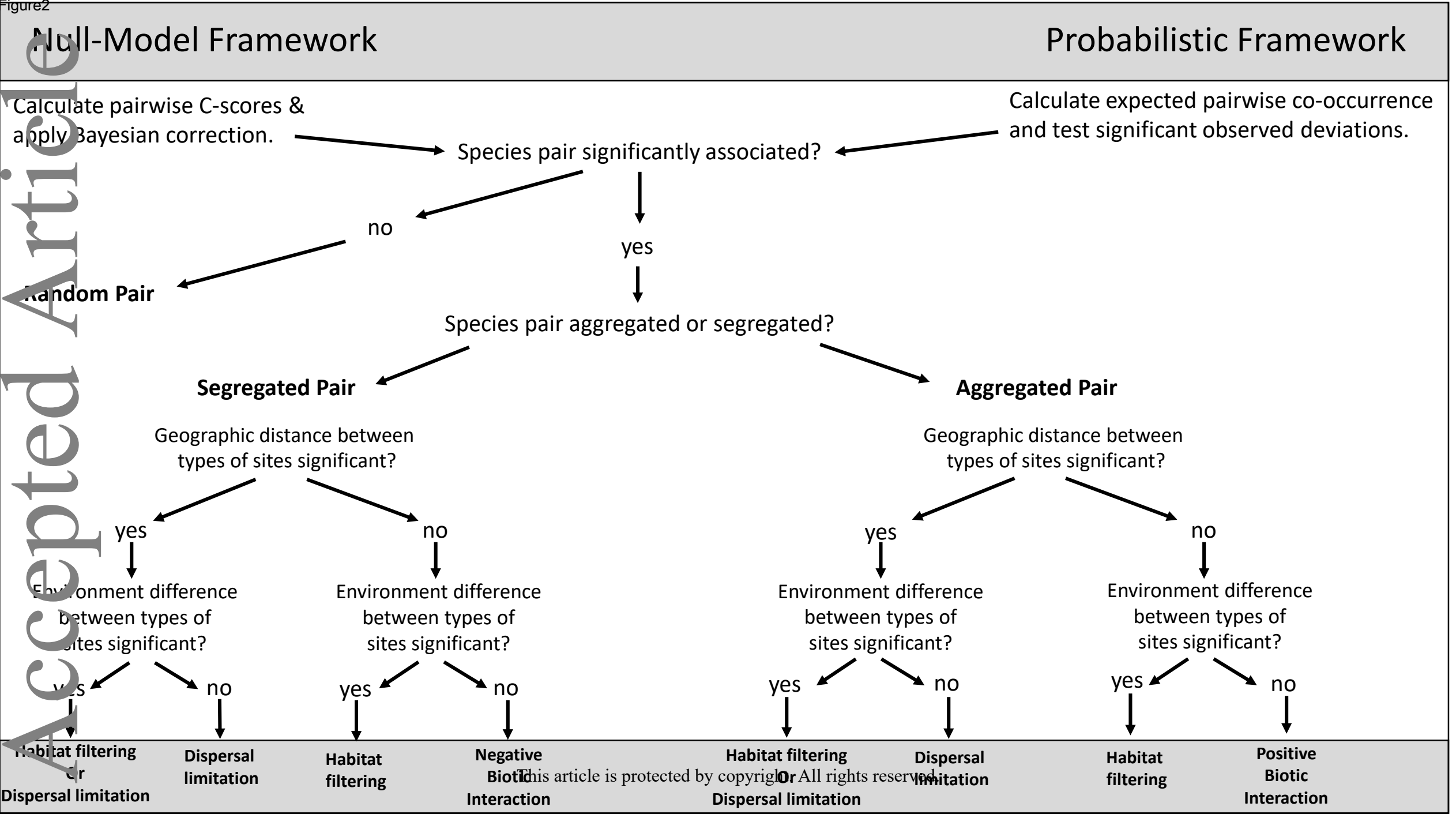
Figure1

Clear Boggy Creek

Muddy Boggy Creek

Oklahoma

Texas

Red River

N

km
0 5 10 20 30 40

Figure2

Null-Model Framework

Probabilistic Framework

Calculate pairwise C-scores & apply Bayesian correction.

Calculate expected pairwise co-occurrence and test significant observed deviations.

Species pair significantly associated?

no

yes

Random Pair

Species pair aggregated or segregated?

Segregated Pair

Aggregated Pair

Geographic distance between types of sites significant?

Geographic distance between types of sites significant?

yes

no

yes

no

Environment difference between types of sites significant?

Environment difference between types of sites significant?

Environment difference between types of sites significant?

Environment difference between types of sites significant?

yes

no

yes

no

yes

no

yes

no

Habitat filtering or Dispersal limitation

Dispersal limitation

Habitat filtering

Negative Biotic Interaction

Habitat filtering or Dispersal limitation

Dispersal limitation

Habitat filtering

Positive Biotic Interaction

Figure3