FINAL REPORT

South Carolina State Wildlife Grant F19AF00720

South Carolina Department of Natural Resources

October 1, 2019 – June 30, 2021

Project Title: Identification of environmental and biological factors limiting occurrence of the Sandhills Chub (*Semotilus lumbee*) in South Carolina

Principal Investigator: Derek P. Crane, Ph.D.
Department of Biology
Coastal Carolina University
dcrane@coastal.edu

Collaborators:
Mark Scott, Ph.D., SCDNR, scottm@dnr.sc.gov

Fritz Rohde, NOAA, Fritz.rohde@noaa.gov

Dustin Smith, North American Native Fishes Association, dsmith73@hotmail.com

Project objectives:

The objectives of this project were to assess Sandhills Chub populations in South Carolina, identify environmental and biological features associated with healthy populations and depleted or extirpated populations, and develop a predictive model to guide Sandhills Chub conservation and restoration.

Project Narrative:

Need

The Sandhills Chub (*Semotilus lumbee*) is categorized as a Highest Priority species under the SC DNR's State Wildlife Action Plan because its global distribution is limited to the Sandhills ecoregion of north-central South Carolina and south-central North Carolina (Bettinger 2005 [revised and edited by Scott et al. 2013]). Additionally, the Sandhills Chub is included on the American Fisheries Society's list of Imperiled Freshwater and Diadromous Fishes of North America (Jelks et al. 2008; accessed online at: https://assets.usgs.gov/fl-biology/afs_fish/map_object.html). Despite being a conservation priority, little research has focused specifically on the Sandhills Chub. Rohde and Arndt (1991) documented the distribution of Sandhills Chub (including new locations and locations where they had been extirpated), and provided descriptive statistics for locations where Sandhills Chubs were located. However, because their objective was to describe the distribution and status of Sandhills Chub, they focused sampling efforts in habitats where Sandhills Chub were likely to be found and thus could not make definitive conclusions about factors leading to population extirpations or declines. Similarly, recent studies investigating fish-habitat relationships in SC streams focused on assemblage-level analyses (Paller et al. 2016; SC Stream Assessment), and therefore do not provide the detailed information necessary to develop a conservation strategy for Sandhills Chub.

As part of the 2015 SC State Wildlife Action Plan, Bettinger (2005 [revised and edited by Scott et al. 2013]) called for research that describes habitat requirements and listed determination of habitat needs as a measure of success for Sandhills Chub conservation. Additionally, Bettinger (2005 [revised and edited by Scott et al. 2013]) recommended resampling known locations and increased sampling in the Sandhills National Wildlife Refuge to assess Sandhills Chub populations. Making comparisons between locations of healthy populations of a species and locations with depleted or extirpated populations can help identify drivers of population loss and potential issues to address with conservation and restoration measures. For example, Rust et al. (2002) used discriminant analysis to identify habitat and water quality characteristics that differentiated lakes with self-sustaining populations of Muskellunge and lakes reliant on stocking. We seek to investigate and identify environmental and biological factors limiting Sandhills Chub populations by comparing habitat and biological characteristics between locations with healthy Sandhills Chub populations and locations with extirpated populations or locations where Sandhills Chub do not occur with healthy populations.

Methods
Sampling was completed in the Sandhills ecoregion portions of the Pee Dee (n = 98) and Wateree (n = 17) watersheds of South Carolina during June through October 2019 and August through November 2020 (Figure 1). Sites were selected using a stratified random sampling design from the population of all wadeable, perennial streams with a drainage area between 2-50 km^2. This drainage area limitation was used because Sandhills Chub are known to be restricted to low-order headwater streams. The two strata used in the sampling design were streams that Sandhills Chub were previously collected in and streams where presence of Sandhills Chub was unknown. We used these two strata to force in sites where Sandhills Chub were present to ensure a sufficient number of events (Sandhills Chub presence) to allow for logistic regression analysis (i.e., fully random sampling may have resulted in too few samples for statistical analysis). This sampling design resulted in a total of 24 sites from streams where Sandhills Chub were previously collected and 91 sites where their status was unknown.

Fish Sampling – At each site, three-pass backpack electrofishing was used to sample a length of 35 times the mean stream width (MSW; Simonson and Lyons 1995). The primary unit used was an ETS Electrofishing Systems model ABP-4 backpack unit (Madison, Wisconsin) with upgraded electrodes (anode: 3/8-inch diameter, cathode: 3 m length) to improve effectiveness in low-conductivity conditions. Due to equipment failure, a second unit (Smith and Root model LR-24, Vancouver, Washington) was used at six sites. The same settings were used at all sites (Pulsed DC, 600V, 25% duty cycle, and a frequency of 40 Hz) regardless of unit used. Despite low conductivity, pilot work and results from this study demonstrated that three-pass electrofishing is effective at detecting Sandhills Chub; Sandhills Chub were detected on the first pass at 36 of the 41 sites where they were collected, on the second pass at five sites and never collected for the first time on the third pass. Therefore, the probability of a type-II detection error using three-pass electrofishing to collect Sandhills Chub was low. We used a standard multi-pass electrofishing technique with block nets at the upstream and downstream end of the reach. Two individuals, both with dip nets and one with the backpack electrofisher, started at the downstream end of each site and sampled by zig-zagging in an upstream direction for three passes. Only two people were used to sample these streams because the average stream width was less than 3 m at most sites, limiting room for any additional individuals in the stream. A single backpack electrofishing unit was used at a time when sampling sites <3m wide. At eight sites, where MSW was >3m, two electrofishing units were used simultaneously with three people sampling. After each pass, all fish were identified, counted, and then released below the downstream block net.

Habitat Sampling – Habitat sampling was completed after fish sampling at each location. Habitat characteristics were collected at the site level using the transect method. Since most of the streams in this study were small (MSW <3 m), 13 evenly-spaced transects were used at each site (Simonson et al. 1994). Width was recorded at each transect and depth was recorded at three points along each transect; one point at the deepest spot in the stream (thalweg) and the other two equally distanced (Simonson 1993). Stream cover was also quantified at each site as present or absent for each transect if any form of instream cover (i.e. undercut bank, log jam, rock, aquatic vegetation, root ball) capable of concealing a fish greater than 10 cm was intersected by the transect. Water quality data (conductivity, dissolved oxygen, pH, and temperature) were collected at each site using a Hach (Hqd) portable multiprobe (Loveland, Colorado). Measurements were taken from the middle of the water column in an area of moderate flow. Substrate data were recorded using the Wolman pebble count method (Wolman 1954). One person zig-zagged from bank to bank stopping 100 times to pick up the piece of substrate nearest their big toe on their right foot and then measured the piece of substrate using a gravelometer along the intermediate-axis. This information was used to calculate d50 as well as the percentage of substrate falling within the 6-11 mm range. We calculated percentage of substrate within the 6-11 mm size class at each site because this is the primary size substrate used by Sandhills Chub to construct their nests (Maurakis et al. 1990).

Impoundments and Watershed Characteristics – To investigate the relationship between stream impoundments and presence of Sandhills Chub, we used Google Earth Pro (7.3.3.7786, Google LLC, Mountain View, CA) to determine the total number of impoundments present in each 12-digit hydrologic unit code (HUC) sampled. First, we imported shape files into Google Earth from the National Hydrography Dataset (USGS 2018a; USGS 2018b) and Watershed Boundary Dataset (USGS et al. 2020) for 12-digit HUC boundaries, NHDwaterbodies, and NHDflowlines. Then, we quantified the number of impoundments visually, using satellite imagery. To prevent potential effects of including ponds not created through impoundment of streams, we only counted impoundments that directly intersected NHDflowlines. Watershed characteristics were analyzed using the National Aquatic Resource Surveys' StreamCat dataset (Hill et al. 2016). Sample locations were linked to the corresponding COMID number and data for percent forested land, percent forested land within 100 m stream buffers, percent of land developed, percent of land developed within 100 m buffers, canal density, road density, road crossing density, percent of land used for crops and hay, and percent of land used for crops and hay within 100m stream buffers were extracted for each watershed. Elevation data was also extracted for the local catchment at each site.

Data Analysis – All analyses were completed using R in RStudio (R Core Team 2021; RStudio Team 2020). First, we used the 'cooccur' package (Griffith et al. 2016) to analyze any potential biological relationships between Sandhills Chub and other species encountered while sampling. Species co-occurrence was calculated using a random sampling with replacement and hypergeometric distribution approach to the probabilistic model (Griffith et al. 2016; Veech 2013) which compares observed co-occurrence to the expected co-occurrence between pairs of species and searches for significant positive, negative, and random pairwise interactions (α = 0.05). This analysis was completed to see if co-occurrence data suggests ecological interactions between Sandhills Chub and other species that may facilitate or prevent their occupation of a site and should therefore be included in the model.

Next, we used a hierarchical generalized linear mixed-model (binary response: present or absent) to investigate the relationships between environmental variables and the probability of presence for Sandhills Chub. Samples were taken at the site level and were nested within 42 different 12-digit HUCs, which were nested within 10, 10-digit HUCs. Therefore, we used 12-digit HUCs nested within 10-digit HUCS as random effect grouping factors. Prior to conducting the analysis, we tested for multicollinearity among variables using the vif() function in the 'car' package (Fox and Weisburg 2019). Because including too many predictor variables relative to the number of events (e.g., locations where Sandhills Chub were present) in a logistic regression model can result in poor model fit and incorrect conclusions (Peduzzi et al. 1996; Vittinghoff and McCullouch 2006; Ranganathan et al. 2017), we investigated if the large number of a priori predictor variables could be reduced to a statistically justifiable number, given the number of sites that had Sandhills chub. Peduzzi et al. (1996) reported that for models with less than 10 events per variable, regression coefficients were biased; however, Vittinghoff and McCullouch (2006) reported that this rule of thumb can be relaxed as long as results are interpreted with caution and more complex models are compared to models containing fewer predictors. Therefore, because our dataset only included 41 sites where Sandhills Chub were present, we needed to reduce the dimensionality of the data and limit the analysis to the most biologically plausible set of predictor variables.

To select the initial suite of variables included in the analysis, we examined co-occurrence plots and notched boxplots of relationships between environmental variables and the presence or absence of Sandhills Chub, and considered this information in the context of what is known about Sandhills Chub ecology and ecology of species it may have negative or positive associations with (Figure 2; Figure 3). In notched box plots, the notches represent the 95% confidence interval of the median and when notches between two plots do not overlap that is good evidence that the medians differ (Chambers 1983). Variables where the notches clearly overlapped were removed from consideration and variables with notches that were close to overlapping were investigated further by considering the range and scale of the units. For example, box plots for depth, velocity, and width were very close to overlapping but when the scale of the units was taken into consideration the median differences were at a very small scale (i.e., approx. 5 cm for depth, 0.03 m/s for velocity, and 0.3 m for width). The same logic was applied to the density of roads within the watersheds where notches did not overlap; however, the difference in medians was only about 0.5 km/km^2 and road density throughout the sample area was low. Additionally, roads were accounted for in the development metric as impervious surfaces and no significant differences were highlighted by examination of the boxplots for watershed wide development or within the 100 m stream buffers. The box plots for conductivity also did not overlap in notches; however, this variable was excluded from the model because there is no evidence of conductivity acting as a limiting factor for fishes; conductivity was generally low throughout the sample area. Next, we fit models for all possible combinations of the suite of biologically relevant variables (including the nested random effects in all models) and ranked them based on AICc and calculated ΔAICc and Akaike weights. Model fit was examined using the 'DHARMa' package (Hartig 2020) to check for overdispersion, underdispersion, and issues with the distribution of residuals, and no significant issues with model fit were identified. Finally, we completed 5-fold cross-validation on the most likely models to calculate the receiver operating characteristic (ROC) curve and estimate the area under the curve (AUC) using the 'pROC' package (Robin et al. 2011). The ROC curve is a plot that shows specificity versus sensitivity for varying levels of a classification threshold in logistic regression. Area under the ROC is a metric used to measure performance of classification models where AUC indicates the probability that the model will rank a true positive higher than a true negative value (i.e., values closer to one indicate more accurate classification).

Results

We captured 7788 fish, representing 52 species, at the 115 sites sampled. The most common species encountered included Dusky Shiner (*Notropis cummingsae*), Bluehead Chub (*Nocomis leptocephalus*), Yellow Bullhead (*Ameriurus natalis*), and Bluegill (*Lepomis macrochirus*), with 431 Sandhills Chubs collected across 41 sites, including 26 sites in streams where Sandhills Chub had not been previously collected. Sampled streams were narrow (mean = 2.58 m wide, SD = 1.04; Table 1), shallow (mean = 21.4 cm deep, SD = 9.1), and had low conductivity (mean = 32.1 µS/cm, SD = 21.7). Overall land-use in the study area was primarily forested with low levels of development and riparian buffers were generally intact; however, the prevalence of impoundments was high (Table 1).

Cooccurrence analysis indicated significant positive relationships between Sandhills Chub and Dollar Sunfish (*Lepomis marginatus*), Creek Chubsucker (*Erimyzon oblongus*), and Margined Madtom (*Notorus insignis*; Table 2). Negative relationships were found between Sandhills Chub and Largemouth Bass (*Micropterus salmoides*), Bluegill, Creek Chub (*Semotilus atromaculatus*), Eastern Mosquitofish (*Gambusia holbrooki*), and Lined Topminnow (*Fundulus lineolatus*; Figure 3). Because examination of these associations did not suggest interactions that facilitated or inhibited presence of Sandhills Chub (see Discussion), fish associations were not included in the logistic regression analysis.

Based on examination of boxplots and consideration of Sandhills Chub biology, variables for instream cover, percent of substrate between 6 and 11 mm, dissolved oxygen, the number of impoundments in the 12-digit HUC, and elevation of the catchment were included in the logistic regression analysis. The top four models based on AICc were within 2ΔAICc, so we interpreted and reported results from all four (Table 3). The four most likely mixed-models indicated that instream cover, dissolved oxygen, and the amount of substrate between 6 and 11 mm were important predictors of Sandhills Chub presence within their range (Table 4). A one unit (mg/L) increase in dissolved oxygen content resulted in an increase in the odds of presence by 74-79%. As instream cover increased by 10%, the odds of presence increased by 34-38%. Finally, a 10% increase in the amount of substrate between 6 and 11 mm resulted in an increase in the odds of Sandhills Chub presence by 101-126%. Although the most likely model included the terms for the number of impoundments within each HUC12 and the elevation of the local catchment, the

95% confidence interval for these odds ratios (Table 4) overlapped zero and therefore the number of impoundments and elevation were not reliable predictors of Sandhills Chub presence.

Conclusion

The Sandhills Chub prefers streams that have high dissolved oxygen content, presence of instream cover, and the substrates necessary to construct their pit-ridge nests. We identified several new locations where Sandhills Chub occur as well as locations with sizable populations. This information can be used to prioritize conservations efforts in the Sandhills by (1) protecting locations with larger populations (2) restoring habitat at sites with low numbers of Sandhills Chub and (3) potential reintroduction of Sandhills Chub to locations containing suitable habitat but where Sandhills Chub have been extirpated. Protection of riparian zones from development in the future and restoration of degraded sites should assist in maintaining cooler water temperatures, which will in turn allow the stream to hold more oxygen, increase recruitment of woody debris for instream cover, and also limit sedimentation of the stream bed. The addition of instream cover such as woody debris for refuge and substrates between the size of 6 and 11 mm for nest construction may help to improve habitat for Sandhills Chub in locations where these habitat features are missing. Our model also allowed us to identify sites with high quality habitat where Sandhills Chub are not present, for example, sites that the models estimated high probability of Sandhills Chub presence but did not contain Sandhills Chub. These locations should be investigated further and if longer reaches within the stream do not document Sandhills Chub, reintroduction into these sites is an option.

Accomplishments

We were able to complete each of our project objectives and through this project we substantially increased our understanding of Sandhills Chub ecology, which will be used to inform conservation efforts for the species. Completion of this project resulted in a master's thesis for CCU graduate student Garrett Herigan. Mr. Herigan successfully defended his thesis in April 2021 and graduated in May. As a graduate student, Mr. Herigan presented the project's findings at the 2021 annual meeting of the Southern Division of the American Fisheries Society. He is currently completing the final draft of a manuscript based on this study, which will be submitted to a peer-reviewed journal by August 2021.

Federal costs and matching funds

The total federal cost for this project was $48,304. Coastal Carolina University provided $28,213 in matching funds through a graduate assistantship to Mr. Herigan during the 2019-2020 academic year, in-kind services from PI Crane, and unrecoverable indirect costs.

Recommendation

We recommend closing the grant

References

Bettinger J. 2005. Sandhills Chub in Supplemental Volume: Species of Conservation Concern, 2015 SC State Wildlife Action Plan. Reviewed and Edited 2013: M. Scott, A. R. Gelder, and M. T. Cribb.

Chambers, J. M., W. S. Cleveland, B. Kleiner, and P. A. Tukey. 1983. Graphical Methods for Data Analysis. Belmont, California: Wadsworth International Group.

Fox, J. and S. Weisberg. 2019. An R Companion to Applied Regression, Third edition. Sage, Thousand Oaks, California, https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Griffith, D. M., J. A. Veech, and C. J. Marsh. 2016. cooccur: Probabilistic Species Co-Occurrence Analysis in R. Journal of Statistical Software. 69(2):1-17.

Hartig, F. 2020. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.3.3.0. https://CRAN.R-project.org/package=DHARMa

Hill, R. A., M. H. Weber, S. G. Leibowitz, A. R. Olsen, and D. J. Thornbrugh, 2016. The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States. Journal of the American Water Resources Association 52:120-128.

Jelks H. L., S. J. Walsh, N. M. Burkhead, S. Contreras-Balderas, E. Diaz-Pardo , D. A. Hendrickson, J. Lyons , N. E. Mandrak, F. Mccormick, J. S. Nelson, et al. 2008. Conservation status of imperiled North American freshwater and diadromous fishes. Fisheries 33:372–407.

Maurakis, E. G., W. S. Woolcott, and J. T. Magee. 1990. Pebble-nests of four *Semotilus* species. Southeastern Fishes Council Proceedings. 22:7-13.

Paller M. H., B. A. Prusha, D. E. Fletcher, E. Kosnicki, S. A Sefick, M. S. Jarrell, S. C. Sterrett, A. M. Grosse, T. D. Tuberville, and J. W. Feminella. 2016. Factors influencing stream fish species composition and functional properties at multiple spatial scales in the sand hills of the southeastern United States. Transactions of the American Fisheries Society 145:545–562.

Peduzzi, P., J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein. 1996. A simulation study of the number of events per variable in logistic regression analysis. Journal of Clinical Epidemiology. 49:1373–1379.

Ranganathan, P., C. S. Pramesh, and R. Aggarwal. 2017. Common pitfalls in statistical analysis: Logistic regression. Perspectives in Clinical Research. 8:148–151.

Robin, X., N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. Sanchez, and M. Müller. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, p. 77.

Rohde, F. C. and R. G. Arndt. 1991. Distribution and status of the Sandhills Chub, *Semotilus lumbee*, and the Pinewoods Darter, *Etheostoma mariae*. The Journal of the Elisha Mitchell Scientific Society. 107:61-70.

Rust, A. J., J. S. Diana, T. L. Margenau, and C. J. Edwards. 2002. Lake characteristics influencing spawning success of Muskellunge in Northern Wisconsin lakes. North American Journal of Fisheries Management 22:834–841.

Simonson, T. D., J. Lyons, and P. D. Kanehl. 1994. Quantifying fish habitat in streams: transect spacing, sample size, and a proposed framework. North American Journal of Fisheries Management 14:607–615.

U.S. Geological Survey. 2018a. USGS National Hydrography Dataset Plus High Resolution (NHDPlus HR) for 4-digit Hydrologic Unit - 0304 (published 20180501): U.S. Geological Survey.

U.S. Geological Survey. 2018b. USGS National Hydrography Dataset Plus High Resolution (NHDPlus HR) for 4-digit Hydrologic Unit - 0305 (published 20180813): U.S. Geological Survey.

U.S. Geological Survey (USGS), U.S. Department of Agriculture - Natural Resource Conservation Service (NRCS), U.S. Environmental Protection Agency (EPA), and Other Federal, State, and local partners (see dataset specific metadata for details ftp://rockyftp.cr.usgs.gov/ngtoc/hydro/outgoing/WBDArchivedMetadata). 2020. USGS Watershed Boundary Dataset (WBD) for 2-digit Hydrologic Unit - 03 (published 20201030).

Veech, J. A. 2013. A probabilistic model for analyzing species co-occurrence: Probabilistic Model. Global Ecology and Biogeography. 22:252-260.

Vittinghoff, E. and C. E. McCulloch. 2007. Relaxing the rule of ten events per variable in logistic and Cox regression. American Journal of Epidemiology. 165:710–718.

Wolman, M. G. 1954. A method of sampling coarse river-bed material. Transactions of the American Geophysical Union 35:951-956.

Table 1. Habitat characteristics for sites in the Carolina Sandhills where Sandhills Chub (*Semotilus lumbee*) were present or absent and overall means. Standard errors provided in parentheses.

Variable	Present	Absent	Overall
Total fish	60.5 (2.0)	71.5 (10.1)	67.7 (7.3)
Fish sepcies richness	8.2 (0.5)	8.4 (0.6)	8.3 (0.4)
Width (m)	2.67 (0.17)	2.53 (0.12)	2.58 (0.10)
Depth (cm)	22.8 (1.4)	20.6 (1.1)	21.4 (0.85)
Water velocity (m/s)	0.11 (0.01)	0.09 (0.01)	0.09 (0.01)
d50	5.48 (0.80)	7.67 (1.30)	6.89 (0.89)
Substrate between 6-11mm (%)	20.2 (2.3)	11.2 (1.3)	14.4 (1.2)
Instream cover (%)	63.4 (2.89)	47.3 (2.90)	53.1 (2.25)
Temperature (ºC)	20.6 (0.6)	21.4 (0.5)	21.1 (0.4)
Dissolved oxygen (mg/L)	8.04 (0.18)	7.12 (0.22)	7.45 (0.16)
Conductivity (µS/cm)	27.53 (3.89)	34.62 (2.26)	32.09 (2.03)
pH	4.79 (0.13)	5.15 (0.10)	5.02 (0.08)
Number of impoundments	45.0 (6.0)	58.5 (4.5)	53.7 (3.6)
Development - watershed (%)	8.32 (1.31)	9.05 (1.02)	8.79 (0.80)
Development - watershed within 100 m buffer (%)	4.54 (0.98)	4.41 (0.63)	4.45 (0.54)
Forested land - watershed (%)	58.08 (1.86)	59.88 (1.57)	59.22 (1.21)
Forested land - watershed within 100 m buffer (%)	79.98 (1.50)	81.50 (1.01)	80.96 (0.84)
Agricultural land - watershed (%)	13.21 (1.55)	15.83 (1.28)	14.90 (1.00)
Agricultural land - watershed within 100 m buffer (%)	3.53 (0.67)	5.30 (0.63)	4.67 (0.47)
Road density - watershed (km/km^2)	1.81 (0.12)	2.13 (0.11)	2.01 (0.09)
Road density - watershed within 100 m buffer (km/km^2)	1.62 (0.13)	1.82 (0.12)	1.74 (0.09)
Road steam crossings - watershed (crossings/km2)	0.42 (0.05)	0.43 (0.04)	0.43 (0.03)
Canal density - watershed (km/km^2)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Elevation - catchment (m)	113.7 (3.3)	101.7 (3.9)	105.9 (2.8)

Table 2. Results of co-occurrence analysis between Sandhills Chub and other species collected while sampling in the Carolina Sandhills in 2019 and 2020. Analysis was completed using the 'cooccur' package in R (Griffith et al. 2016). Species co-occurrence was calculated using a random sampling with replacement and hypergeometric distribution approach to the probabilistic model (Griffith et al. 2016; Veech 2013) which compares observed co-occurrence to the expected co-occurrence between pairs of species and searches for significant positive, negative, and random pairwise interactions ($\alpha = 0.05$). *P* corresponds to the probability that the species co-occur more or less than what would be expected by random chance and can be interpreted as a *P* -value.

Species	Observed cooccur	Expected cooccur	Relationship	*P*-value
Dollar Sunfish (*Lepomis marginatus*)	29	23.4	(+)	0.023
Creek Chubsucker (*Erimyzon oblongus*)	22	16.7	(+)	0.029
Margined Madtom (*Notorus insignis*)	22	16.4	(+)	0.020
Bluegill (*Lepomis macrochirus*)	10	18.2	(-)	0.001
Largemouth Bass (*Micropterus salmoides*)	9	15.6	(-)	0.006
Eastern Mosquitofish (*Gambusia holbrooki*)	7	16	(-)	0.000
Creek Chub (*Semotilus atromaculatus*)	1	4.8	(-)	0.016
Lined Topminnow (*Fundulus lineolatus*)	0	3.3	(-)	0.012

Table 3. Summary table of the top ten models fit to analyze habitat characteristics associated with presence of Sandhills Chub (*Semotilus lumbee*). DO = dissolved oxygen (mg/L), substrate = the percent of substrates in the 6-11 mm size class, cover = the percent of instream cover, impoundments = the number of stream impoundments within each 12-digit HUC, elevation = the elevation of the local catchment (m).

Predictors	AICc	ΔAICc	Akaike weight
cover + DO + substrate + impoundments + elevation + (1|HUC10:HUC12)	119.95	0	0.23
cover + DO + substrate + impoundments + (1|HUC10:HUC12)	120.62	0.67	0.17
cover + DO + substrate + (1|HUC10:HUC12)	120.65	0.7	0.16
cover + DO + substrate + elevation + (1|HUC10:HUC12)	120.82	0.87	0.15
DO + substrate + impoundments + elevation + (1|HUC10:HUC12)	122.66	2.71	0.06
DO + substrate + impoundments + (1|HUC10:HUC12)	123.15	3.2	0.05
DO + substrate + elevation + (1|HUC10:HUC12)	123.72	3.77	0.04
DO + substrate + (1|HUC10:HUC12)	123.76	3.81	0.03
cover + DO + elevation + (1|HUC10:HUC12)	125.31	5.36	0.02
cover + DO + impoundments + elevation + (1|HUC10:HUC12)	125.43	5.48	0.02

Table 4. Summary of the top four logistic regression mixed-models used to analyze habitat characteristics associated with presence of Sandhills Chub (*Semotilus lumbee*). Odds ratios and 5-fold cross validation values are reported with 95% confidence intervals in parentheses. Odds ratios for cover, substrate, impoundments, and elevation are according to a 10-unit increase and the odds ratio for DO is according to a 1-unit increase. DO = dissolved oxygen (mg/L), substrate = the percent of substrates in the 6-11 mm size class, cover = the percent of instream cover, impoundments = the number of stream impoundments within each 12-digit HUC, elevation = the elevation of the local catchment (m).

Model Rank	1	2	3	4
Cover	1.34 (1.02-1.79)	1.36 (1.02-1.81)	1.38 (1.04-1.84)	1.36 (1.02-1.82)
DO	1.75 (1.10-2.78)	1.79 (1.12-2.88)	1.76 (1.09-2.83)	1.74 (1.09-2.77)
Substrate	2.08 (1.16-3.73)	2.26 (1.25-4.09)	2.20 (1.20-4.03)	2.01 (1.10-3.65)
Impoundments	0.81 (0.64-1.04)	0.84 (0.67-1.06)	-	-
Elevation	1.21 (0.90-1.63)	-	-	1.14 (0.85-1.53)
5-fold AUC	0.79 (0.60-0.97)	0.79 (0.61-0.96)	0.81 (0.62-0.98)	0.81 (0.62-0.98)

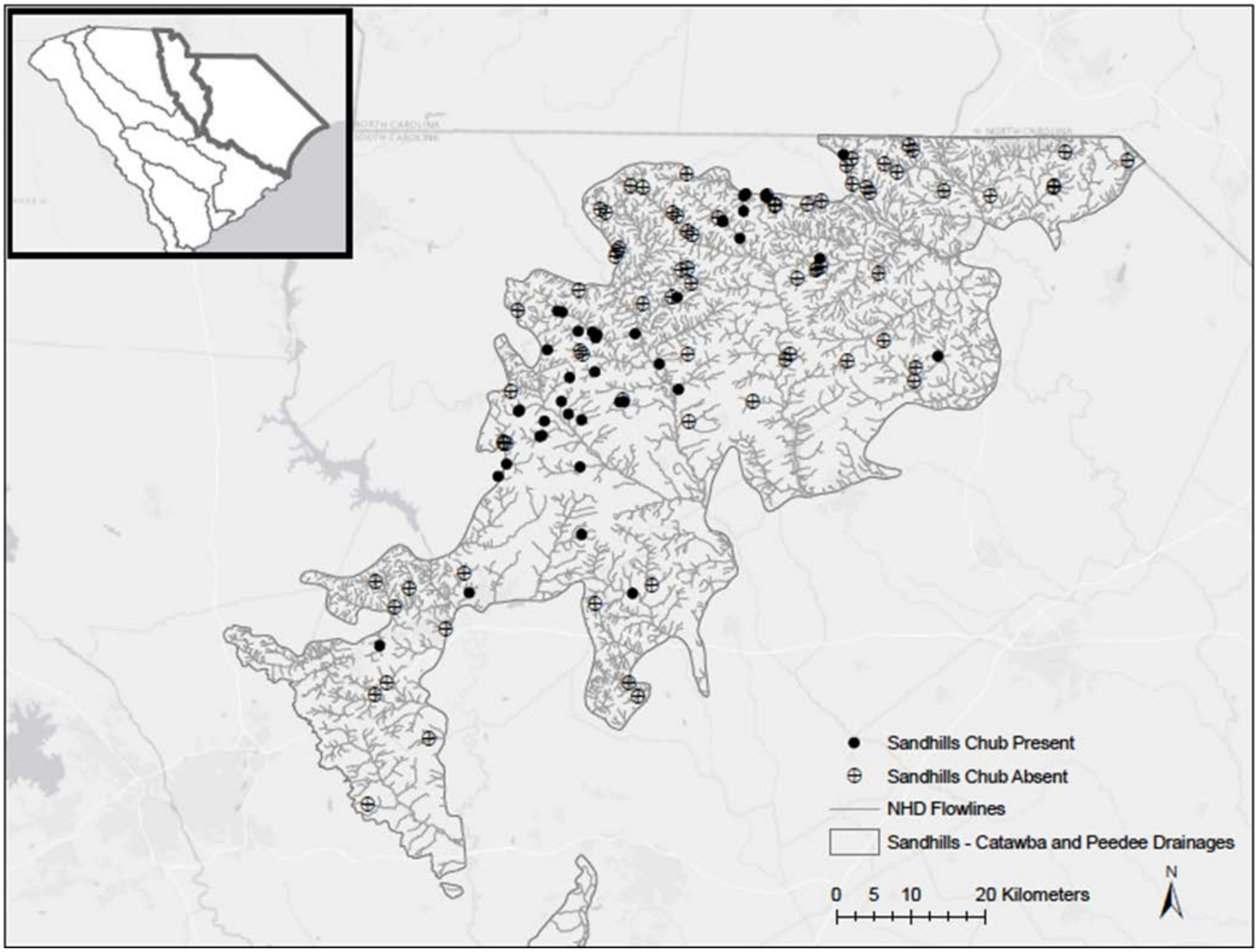


Figure 1. Map of the Sandhills ecoregion of South Carolina within the Catawba and Pee Dee river basins. Solid dots show sites where Sandhills Chub (*Semotilus lumbee*) were present and open circles with a plus show sites where they were absent.

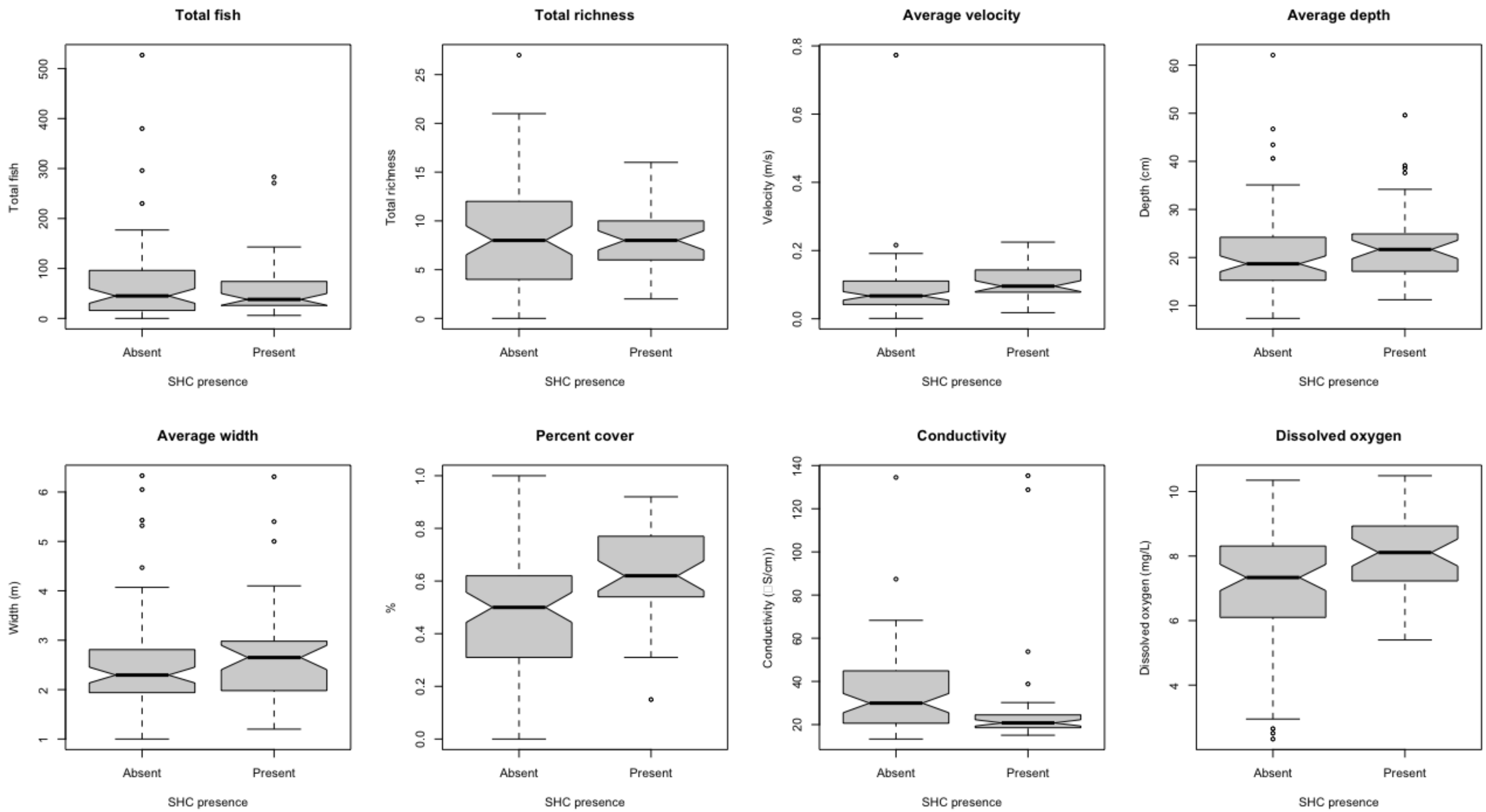
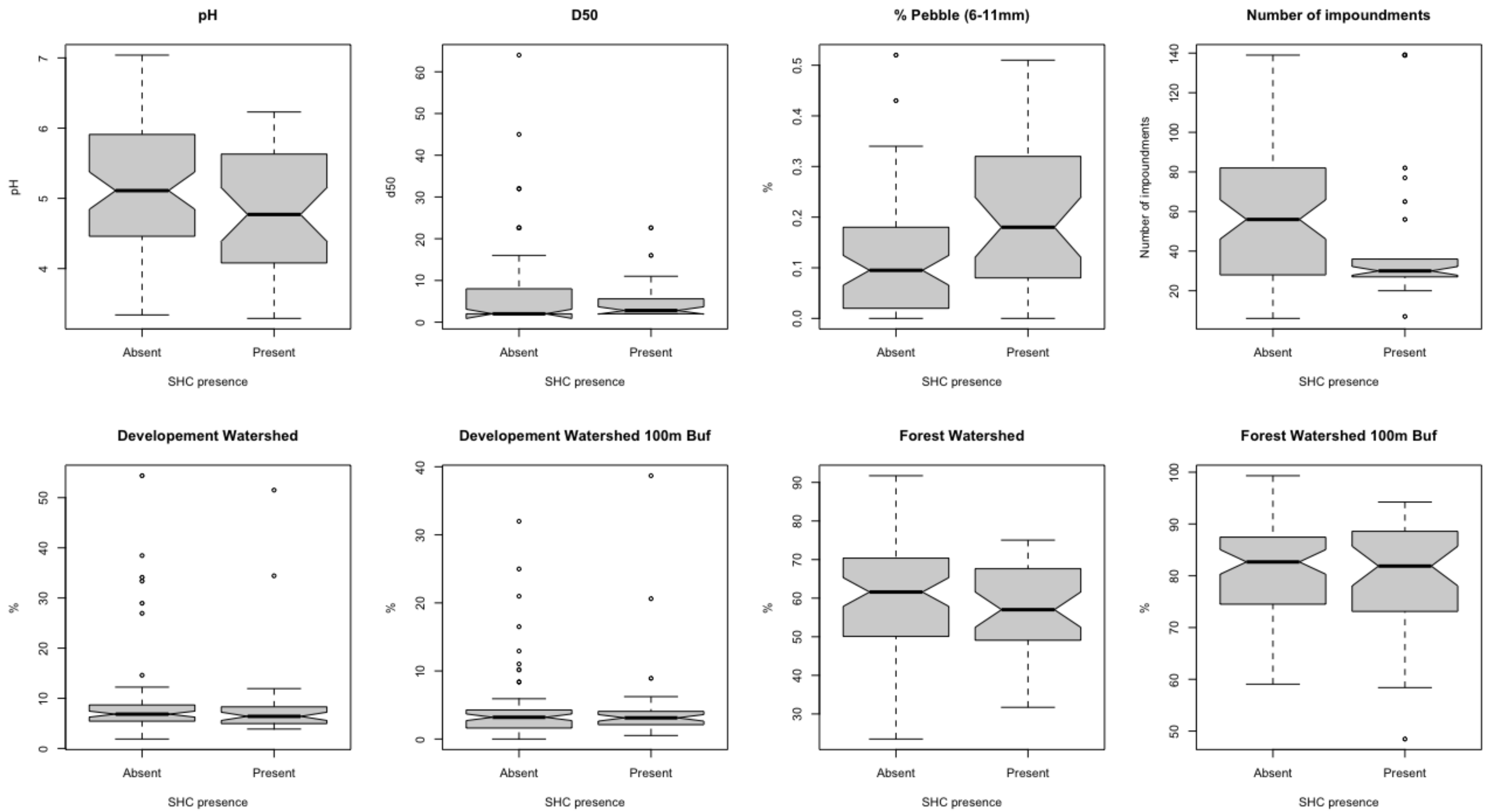
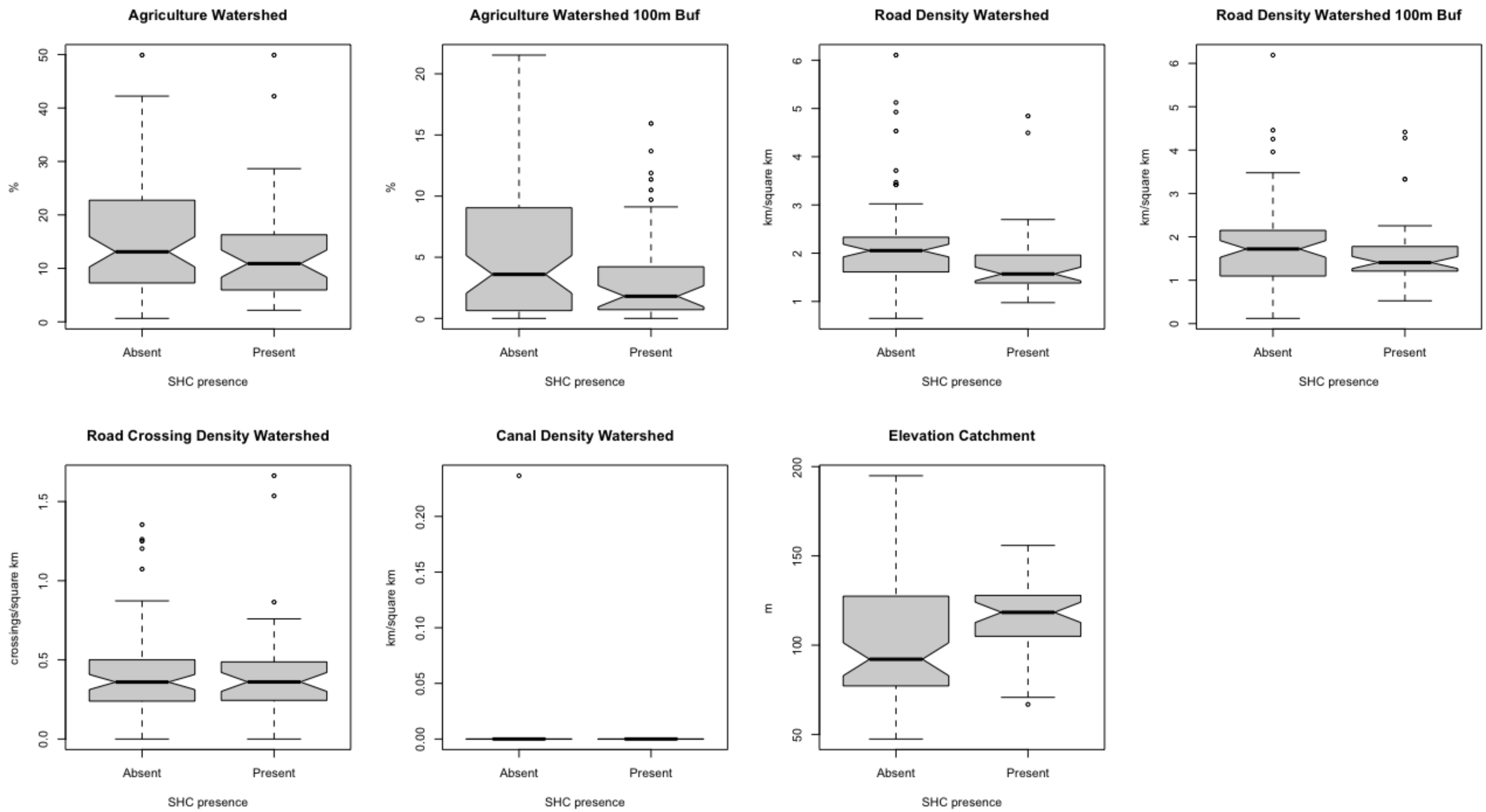


Figure 2. Notched boxplots showing differences between habitat characteristics at sites where Sandhills Chub (*Semotilus lumbee*) were present vs. absent. Notches represent 95% confidence interval of the median.



pH
D50
% Pebble (6-11mm)
Number of impoundments
Developement Watershed
Developement Watershed 100m Buf
Forest Watershed
Forest Watershed 100m Buf
pH
d50
%
Number of impoundments
Absent
Present
SHC presence


Agriculture Watershed
%
Absent
Present
SHC presence
Agriculture Watershed 100m Buf
%
Absent
Present
SHC presence
Road Density Watershed
km/square km
Absent
Present
SHC presence
Road Density Watershed 100m Buf
km/square km
Absent
Present
SHC presence
Road Crossing Density Watershed
crossings/square km
Absent
Present
SHC presence
Canal Density Watershed
km/square km
Absent
Present
SHC presence
Elevation Catchment
m
Absent
Present
SHC presence

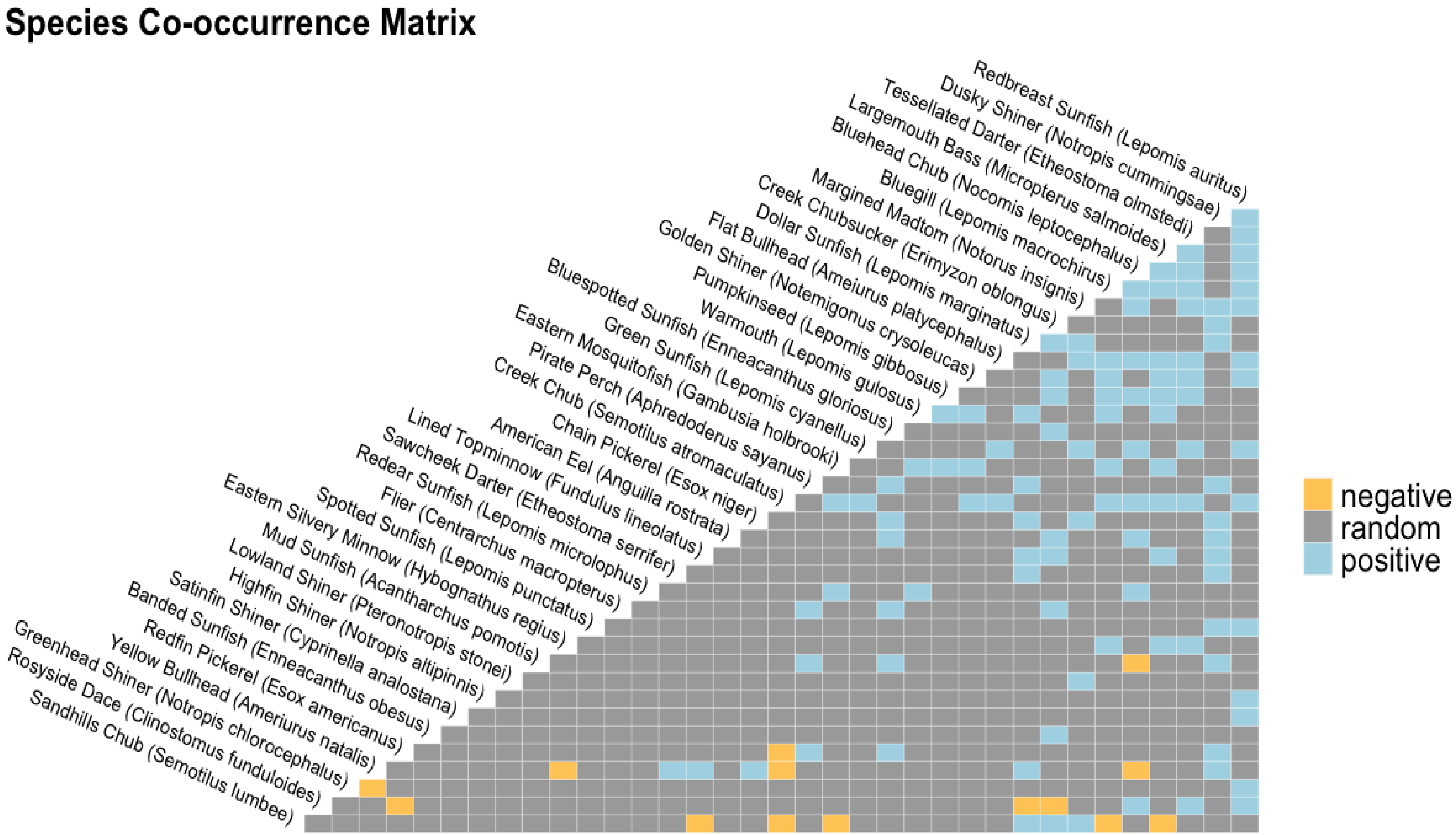


Figure 3. Plot of co-occurrence analysis showing the relationship between Sandhills Chub (*Semotilus lumbee*) and other fish species captured in the Carolina Sandhills during sampling in 2019 and 2020. Species co-occurrence was calculated using a random sampling with replacement and hypergeometric distribution approach to the probabilistic model (Griffith et al. 2016; Veech 2013) which compares observed co-occurrence to the expected co-occurrence between pairs of species and searches for significant positive, negative, and random pairwise interactions ($\alpha = 0.05$).