2021

Environmental Controls on the Spatial Distribution of Greenfin Darters and Biodiversity in the Blue Ridge Mountains

Dri Tattersfield

Recommended Citation

Tattersfield, Dri, "Environmental Controls on the Spatial Distribution of Greenfin Darters and Biodiversity in the Blue Ridge Mountains" (2021). *CMC Senior Theses*. 2591.
https://scholarship.claremont.edu/cmc_theses/2591

Environmental Controls on the Spatial Distribution of Greenfin Darters and Biodiversity in the Blue Ridge Mountains

A thesis presented by

Dri Tattersfield

To the Keck Science Department

of the Claremont McKenna, Scripps and Pitzer Colleges

in partial fulfillment of

the degree of Bachelor of Arts

Senior Thesis in Physics

November 23, 2020

Table of Contents

Acknowledgements

Everything happens through collaboration, especially science. This thesis would not have been possible without the patient and insightful guidance of Maya Stokes, the best mentor over the past several months that I could have asked for, or the wisdom of my advisors Taylor Perron and Branwen Williams. To Dr. Perron and Maya, thank you so much for taking me on this summer amidst the chaos of the pandemic and welcoming me into the Perron Lab. To Dr. Williams, thank you for your extraordinary support and feedback throughout this semester (and last semester, too!). I have grown so much as a scientist and as a person this year thanks to you all.

Thank you to the members of the Perron Lab and Aviva Intveld not only for giving me helpful feedback throughout the summer and teaching me a *lot* about earth science, but also showing me what community in science looks like. Thank you to Sarah Woo, Dayla Woller and Ava McIlvaine for the solidarity and jokes throughout the fall semester. Additionally, thank you to the MIT Summer Research Program, the CMC Soll Center for Student Opportunity, and the Beckwiths for making my summer research experience possible.

Thank you to Stone Van Allen, Madison Yardumian, and Keila Fisher for listening to me constantly talk about fish and for creating the most lovely home this fall. Mackenzie Priest-Heck, Laleh Ahmad, Dina Rosin and James Slaughter, you have kept me grounded in a year where that feels like an impossible feat.

Finally, to Mom, Dad, Jaclyn and Evan, I am always grateful.

Abstract

Disproportionate concentrations of biodiversity in mountains worldwide suggest linkages between geologic processes and biodiversity that are not yet well understood. The Tennessee River Basin in the Blue Ridge Mountains of the southeastern U.S. is a global hotspot for freshwater fish biodiversity. To investigate drivers of biodiversity in the Tennessee River Basin, and explore links to geologic processes, I study the Greenfin Darter (*Nothonotus chlorobranchius*), a small fish endemic to the upper Tennessee River Basin. I use generalized linear models (GLMs) to evaluate the influence of topography, lithology, climate and land use on the distribution of the Greenfin Darter, and find that slope, elevation, geologic age, soil erosion, temperature and pasture cover drive where Greenfin Darters live. Next, I conduct additional topographic and genomic analysis to examine the hypothesis that steps in topography, or knickpoints, isolate Greenfin Darters and lead to genetic divergence. I find tentative evidence that knickpoints may play a role in geographically isolating Greenfin Darter populations and causing allopatric speciation. Finally, I analyze spatial correlations between freshwater fish species richness and anthropogenic environmental impacts and find a weak negative correlation between Superfund sites and darter species richness in the southeastern U.S. These results highlight that the unique biodiversity of the Tennessee River Basin may be at risk from climate and land use change. Furthermore, these results suggest that topographic and lithologic variation may contribute to biodiversity by creating ecological niches and causing speciation, a starting point for understanding how geologic processes shape biodiversity and evolution in mountains globally.

Key words: Greenfin Darter, biodiversity, biogeography, species distribution modeling, knickpoints, speciation, Tennessee River Basin

Introduction

The Greenfin Darter (*Nothonotus chlorobranchius*) is a small fish endemic to the upper Tennessee River Basin in the Blue Ridge Mountains of the greater Appalachian Mountain range (Figure 1). This area hosts the highest freshwater fish species richness and highest concentration of endemic fish species across the entire continental United States; it is home to more fish species than the entirety of Europe (Etnier and Starnes, 1993). The Blue Ridge Mountains are also a site of historical and ongoing displacement of indigenous communities including the Cherokee Nation and Eastern Band of Cherokee Indians (ebci.com), and span vast wealth disparities (Kolmar, 2018). This project explores the biogeography of the Greenfin Darter in the Blue Ridge Mountains by characterizing the distribution of the Greenfin Darter, and interpreting the mechanisms through which geologic, climatic, and anthropogenic processes shape this distribution. Finally, I discuss what we can learn from the distribution of the Greenfin Darter about the relationships between these processes and biodiversity.

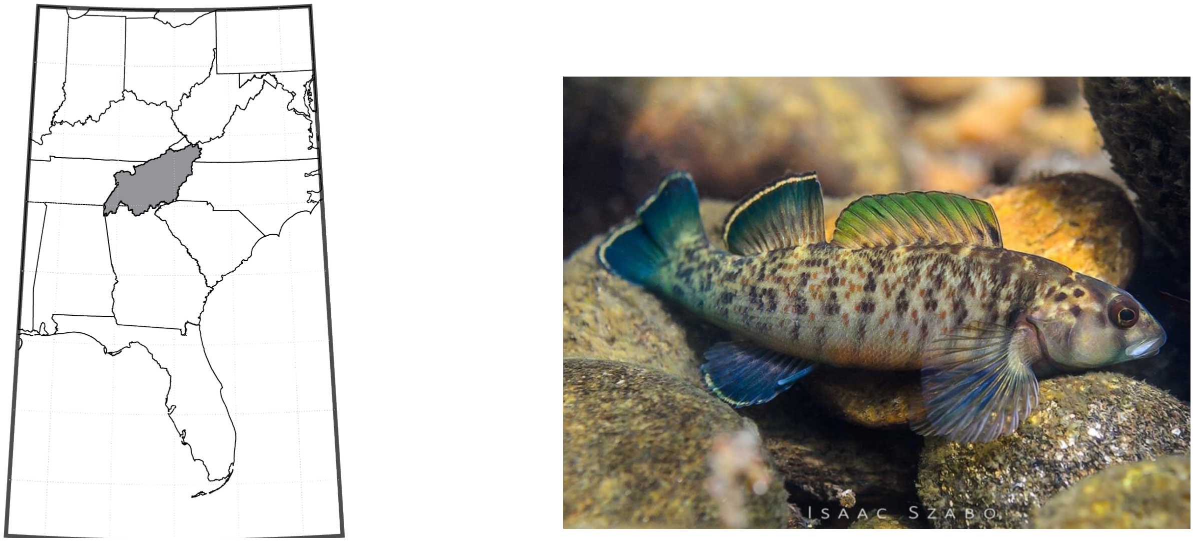


Figure 1a: The location of the Upper Tennessee River Basin (shaded gray) within the continental United States. Figure 1b: The Greenfin Darter (Szabo)

A secondary focus of this project is understanding links between geologic processes and biodiversity. I do this by exploring whether knickpoints, which are steps in topography that often occur across rock type boundaries, 1) isolate Greenfin Darters upstream, explaining their current distribution, and 2) isolate Greenfin Darter populations from each other, causing allopatric speciation and contributing to regional biodiversity.

The project uses Geographic Information Systems (GIS) mapping to visualize where the Greenfin Darter lives in the context of spatial patterns in precipitation, temperature, elevation and other environmental variables. Then, I quantify these relationships using species distribution modeling to identify the most influential factors constituting suitable Greenfin Darter habitat using 283 geologic, climatic, hydrologic and anthropogenic variables.

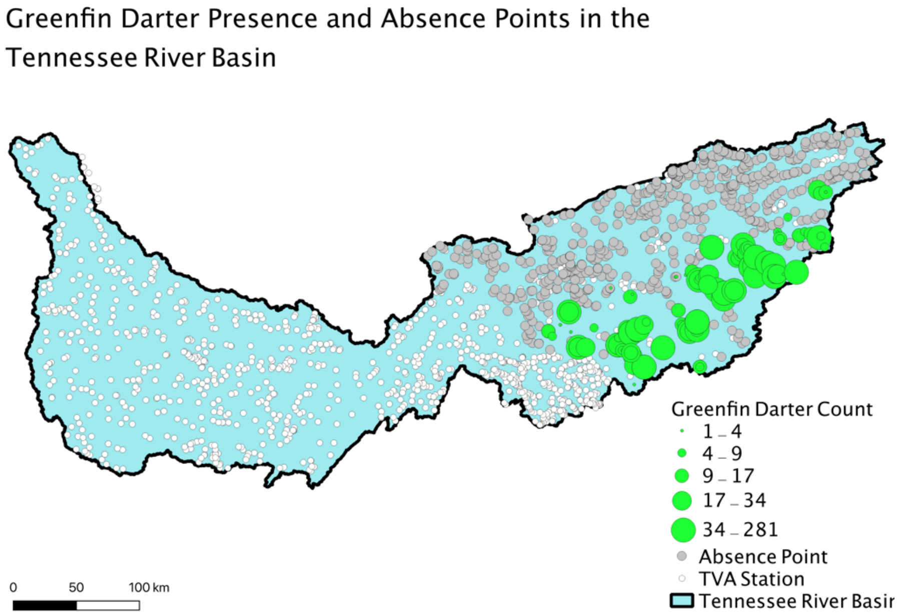


Figure 2: An outline of the full Tennessee River Basin (Hydrologic Unit Code 2-06). Green points, size representing count, show Greenfin Darter presence points used in this project. Gray points show absence points used in this project. The white points represent TVA stations in the lower Tennessee River Basin and points overlapping the range of *Nothonotus camurus* excluded from modeling.

From the species distribution model, I investigate four hypotheses about what factors influence where the Greenfin Darter lives in the upper Tennessee River Basin. The locations of Greenfin Darter populations are shown in Figure 2.

1. Topography: knickpoints, steps in topography along the river, prevent fish movement and thus isolate Greenfin Darter populations upstream.
2. Lithology: The Greenfin Darter prefers to live on older, metamorphic rock.
3. Climate: The Greenfin Darter prefers to live in colder areas.
4. Land use: The Greenfin Darter prefers to live away from pollution and human activity.

As I evaluate these hypotheses about Greenfin Darter habitat, I also analyze the potential role of these processes in shaping freshwater fish species richness in the Tennessee River Basin. In particular, I conduct topographic and genomic analysis to examine the role of knickpoints in determining Greenfin Darter distribution, and evaluate whether they may play a role in driving speciation. Additionally, I compare maps of freshwater fish species richness to dam locations and pollution patterns to explore how human activity relates to biodiversity.

Background

Mountains occupy only 25% of global land area but host over 85% of the world's species (Rahbek et al, 2019). This pattern, also known as Humboldt's Enigma, may be linked to the topographic complexity of mountain regions and their associated unique and heterogeneous climatic conditions (Rahbek et al, 2019). However, the relationships between geologic processes, physical landscape characteristics, and biodiversity are not well understood. In the Tennessee River Basin, variation in rock type creates erosion patterns that cause the formation of knickpoints, steep sections of the river, and drainage basin contraction and expansion (Gallen, 2018). These changes in river network topology may influence the evolution, speciation and dispersal of aquatic species, increasing biodiversity (Gallen, 2018; Albert et al., 2018; Stokes & Perron, 2020).

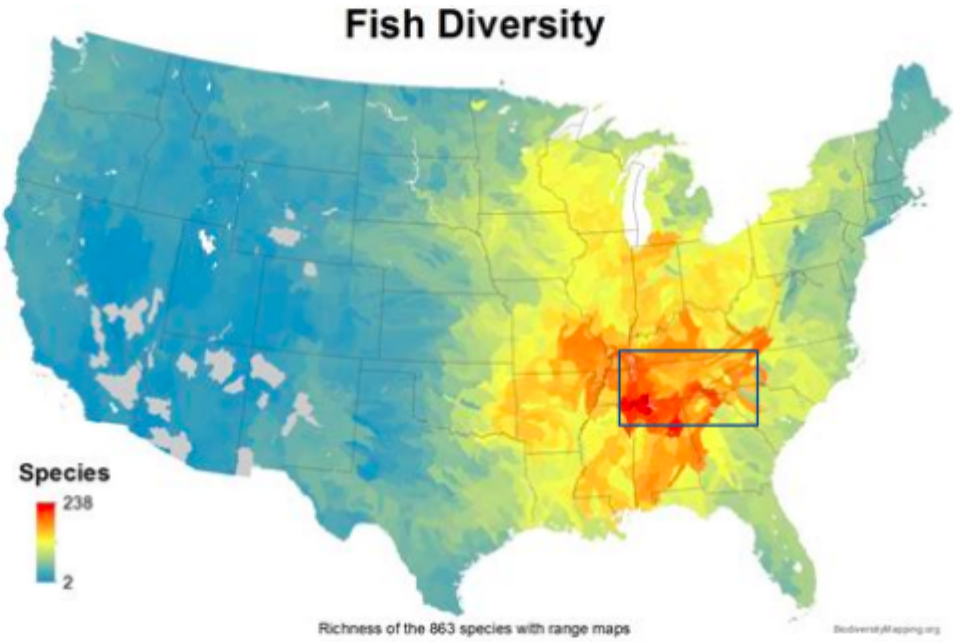


Figure 3: Species richness of freshwater fish in the United States, with a box showing the location of the Tennessee River Basin. (Jenkins et al, 2015)

However, the high species richness of the Tennessee River Basin is at risk. Most endemism occurs in the high-elevation streams, where species have adapted to cool, clear conditions, while generalist species occupy warmer, fine-sediment-rich lowland waters (Scott & Helfman, 2011). Thus, as temperatures rise due to climate change, and land use practices change sediment conditions, fish assemblages may become more homogenized if specialist endemic species cannot adapt to changing conditions (Scott & Helfman, 2011). Thus, understanding the habitat conditions of highland fish such as darters is not only of interest in understanding why the

Tennessee River Basin has such high biodiversity, but also in understanding how that biodiversity may change in the future.

Deforestation and urban development measures such as building and road density have been found to be stronger predictors than topographic features for endemic highland fish species across the southeastern U.S. (Scott, 2006). Furthermore, a study of fish assemblages in the French Broad tributary of the Tennessee River Basin found that agricultural land cover was the primary driver of fish assemblage composition, and secondary drivers were urban land cover, metal concentration, and soil erodibility (Rashleigh, 2004). However, studies specifically focused on the Greenfin Darter have not been conducted.

Greenfin Darters inhabit fast-moving, rocky creeks and small to medium rivers (Fishbase). Mapping the Greenfin Darter's distribution shows that it lives in the high-elevation areas of the Upper Tennessee River Basin, in cooler, wetter regions. Additionally, Figures 4a-4d show that the presence of metamorphic rock, higher elevation, higher precipitation and lower temperatures are spatially congruent.

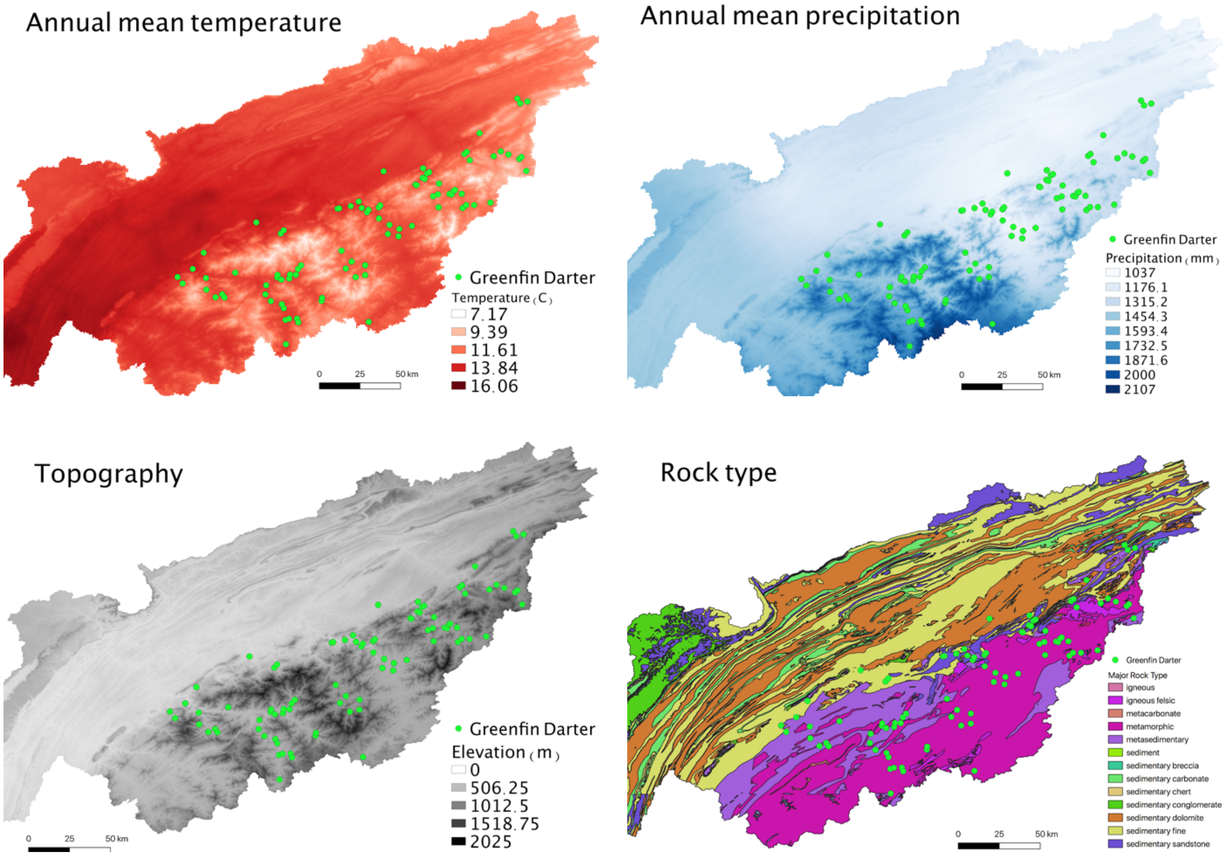


Figure 4: Greenfin Darter localities mapped over a) annual mean precipitation (mm), b) annual mean temperature (C), c) elevation (m) and d) rock type.

This thesis was conducted within the context of a larger project led by PhD student Maya Stokes and Dr. Taylor Perron studying the relationship between knickpoints, rock-type and biodiversity in the Tennessee River Basin. The central hypothesis of Stokes' project is that knickpoints isolate populations from each other, resulting in allopatric divergence and speciation (Stokes et al, 2019). Shown in Figures 5a and 5b, the topologies of the phylogenetic tree of the Greenfin Darter and the river network structure of the upper Tennessee River Basin are closely correlated. In order to travel between tributaries, individuals would be required to travel over knickpoints and different rock-types. The phylogenetic data thus suggests a role for either topography or geology in driving genetic divergence in the Greenfin Darter. Here, I will assess the role of these factors in controlling the modern distribution of the Greenfin Darter in order to provide insight into links between geology, topography and biological evolution.

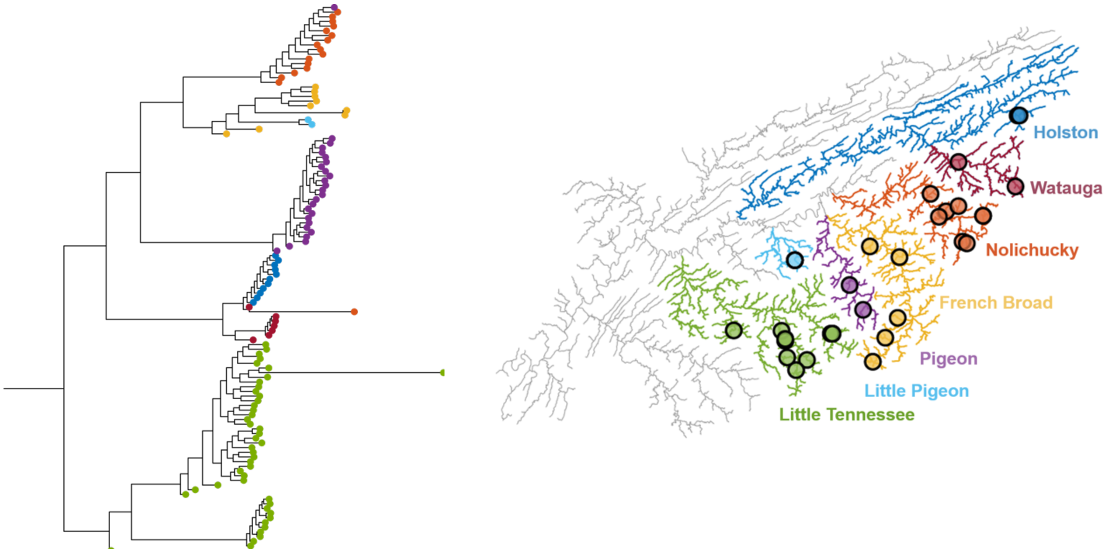


Figure 5a: A phylogenetic tree of the Greenfin Darter, populations colored by the tributary in which they reside. Figure 5b: The Greenfin Darter populations in Figure 5a mapped onto their locations in the Upper Tennessee River Basin (Stokes et al, 2019)

While the larger project investigates how darter ranges and the landscape have evolved over time, my thesis focuses on the state of the landscape and darter habitat in the present. In context of the overall project, the goal of my work is to determine the relative influence of knickpoints and other environmental variables on the range of the Greenfin Darter, through my four hypotheses: topography, lithology, climate, and anthropogenic land use.

Methods

To begin understanding the distribution of the Greenfin Darter, I used QGIS to create maps of Greenfin Darter locations and landscape variables, using projection UTM 17N (QGIS.org, 2020). Then, I used species distribution models with a wide range of environmental predictors to investigate factors shaping the Greenfin Darter's distribution across the Tennessee River Basin. To further evaluate my hypotheses and make connections to biodiversity overall, I conducted additional analysis to explore the role of topography in both Greenfin Darter distribution and speciation across the Tennessee River Basin in more detail. Finally, in my discussion, I analyzed spatial relationships between darter abundance, freshwater fish species richness and anthropogenic and socioeconomic variables.

Data

The Greenfin Darter distribution data came from the Tennessee Valley Authority (TVA), which monitors biodiversity in the Tennessee River Basin by conducting species counts at 1758 stations throughout the basin (Jeff Simmons, personal communication). For each station, I calculated the mean Greenfin Darter abundance across all of the counts that have been conducted at that station since 1990. Because the stations are spread extensively and evenly across the basin, I had reliable absence data as well as presence/abundance data for the Greenfin Darter. I excluded absence points within 7 km of a presence point to minimize noise; if individuals have been found at a location, they are likely to be present along that reach of the river. Additionally, I excluded absence points downstream of the Hiwassee tributary, where the range of *Nothonotus camurus* begins (Figure 2). Within the range of *N. camurus*, the absence of the Greenfin Darter may be more likely to be due to ecological interactions between the two species rather than landscape variables and habitat suitability. The presence and absence points used in the model as well as the locations of the rest of the TVA stations are shown in Figure 2.

For our environmental predictors, I used the RiverATLAS dataset and state geologic maps. The RiverATLAS dataset is a collection of data from different global models, including 56 variables and 281 total attributes in the categories of hydrology, physiography, climate, land cover, soils & geology, and anthropogenic (Linke et al, 2019) (Appendix A). The only modification I made to this dataset was to replace temperature and precipitation data from the WorldClim v1.4 dataset, which uses climate records from 1950-2000, with data from the WorldClim v2.1 dataset, which uses climate records from 1970-2000 (Fick and Hijmans, 2017). Geologic map data is from the state geologic maps of North Carolina, Tennessee, Virginia and Georgia from the USGS national map compilation (Horton, 2017). These maps include age estimates as well as major rock-type, both of which I incorporate into my species distribution models.

To explore the relationship between topography and speciation, I used phylogenetic data presented collected and assembled by Professor Thomas Near and his research group at Yale University, as well as the Tennessee Valley Authority (Figure 5). The tissue samples used in this study are housed at the Yale Peabody Museum Fish Tissue Collection. Maya Stokes, Edgar Benavides and Daemin Kim conducted the laboratory work to build genomic libraries for the Greenfin Darter. The genomes were sequenced with a method called radSEQ which sparsely samples single nucleotide polymorphisms (SNP's) across the entire genome. The genetic distance data used in this paper was derived from the radSEQ data. The value they measured is called D_{XY}; it is a measure of the number of shared nucleotides divided by the length of the genome. This data was used to analyze genetic distances between Greenfin Darter pairs in different locations relative to knickpoints in the Upper Tennessee River Basin.

Finally, I conduct additional analysis on anthropogenic impacts on biodiversity. I use freshwater fish species richness data from NatureServe, which shows numbers of freshwater fish species in each small watershed across the continental United States, delineated by 8-digit Hydrologic Unit Codes from the USGS (NatureServe, 2010). This dataset also divides species by taxonomic classifications (NatureServe, 2010). I use data on dam locations throughout the continental United States from the U.S Department of Transportation (Rawson, 2016). As a proxy for point source industrial pollution, I use the locations of Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) sites, informally known as Superfund sites. These sites include manufacturing facilities, processing plants, landfills and mining sites in which hazardous waste has been improperly managed (U.S. EPA, 2017). Data for Superfund site locations comes from the Superfund National Priorities List Where You Live Map (U.S. EPA, 2015).

Species Distribution Modeling

Species distribution models use environmental data to predict the distribution of species throughout space. They are used to both predict unknown species ranges and understand environmental variables shaping species ranges. The most common species distribution models for studies of marine environments are Maxent, a machine learning approach, Generalized Additive Models (GAMs) and Generalized Linear Models (GLMs) (Melo-Marino et al, 2020). A GLM is a form of linear regression that accepts distributions other than normal distributions. I initially tested both the Maxent and GLM approaches, and ultimately chose the GLM because it can accept both presence-only and presence-absence data, while Maxent only accepts presence data.

I used MATLAB 2020a for all modeling and analysis (Mathworks). The modeling process had three stages. First, I ran the GLM, using the *fitglm* function, using all RiverATLAS variables, rock type and geologic age data, and the darter abundance and absence data. Second, to reduce the number of variables, I grouped variables by their category according to the RiverATLAS classification, and ran the GLM on each group. These smaller model runs allowed us to identify significant variables ($p < 0.05$) from each category to include in the final model. Finally, I ran the GLM using only these final variables. I evaluated the model by training it on 70% of the abundance-absence data and testing its ability to correctly predict the remaining 30% of the data, using the *predict* function in MATLAB.

Topographic Analysis

Next, I further examined my first hypothesis, that the range of the Greenfin Darter is influenced by knickpoints. The river network and corresponding elevation was derived from the HydroSHEDS digital elevation model (DEM) (Lehner, 2008) using the TopoToolbox 2 library in MATLAB, which I used for all topographic analysis (Schwanghart, 2014). First, I located knickpoints along each tributary of the river basin by mapping the river profile using the *plotdz* function, and visually identifying the steepest section along the main trunk of each tributary (Figure 7a).

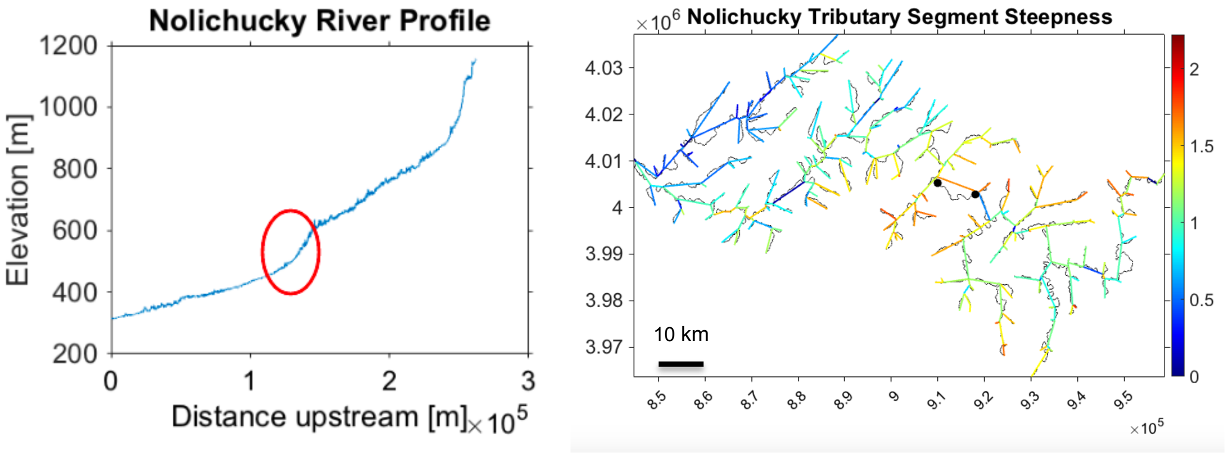


Figure 7a: The river profile used to identify the knickpoint in the Nolichucky tributary of the Tennessee River Basin. The steep section circled in red is where we identified the knickpoint. Figure 7b: A plot representing mean steepness along each segment of the Nolichucky tributary. The black points represent the ends of our identified knickpoint.

To verify my selections, I divided each tributary into segments and mapped the normalized river steepness, or ksn, along each segment, using the *plotsegmentgeometry* function. Normalized river steepness is a common topographic metric in geomorphology that normalizes the stream gradient by drainage area (Wobus, 2006). I considered a knickpoint correctly identified when the segments between the ends of the knickpoint were steeper than surrounding segments (Figure 7b). Once I identified knickpoints throughout the upper Tennessee River Basin, I analyzed relationships between knickpoint steepness and Greenfin Darter distribution using the *gradient* function.

Anthropogenic impact analysis

Finally, I conducted additional analysis on anthropogenic impacts on Greenfin Darter distribution as well as biodiversity in the region overall. I used QGIS to map the locations of dams and Superfund sites with Greenfin Darter distributions and freshwater fish species richness data across the southeastern United States. In order to quantify relationships between dams, Superfund sites and biodiversity, I calculated spatial statistics by using the *Join attributes by location (summary)* function in QGIS. Then, I calculated correlation coefficients between species richness and number of dams and Superfund sites using the *corrcoef* function in MATLAB.

Results

Species Distribution Modeling

The initial model run with all environmental variables had a rank deficient regression matrix due to there being significantly more variables (283) than darter presence locations (89). In the second modeling stage, these variables were separated into categories, and a GLM was run for each category. The significant variables ($p < 0.05$) from each category, 21 in total, were entered into a GLM together for the final model. These were geologic age, slope, elevation, soil erosion, soil water content, temperature, precipitation, potential and actual evapotranspiration, snow cover, pasture cover, runoff, gross domestic product (GDP), human development index (HDI), groundwater table depth, discharge, and the categorical rock type variable. Full model results for each of these model runs are shown in Appendix B. These variables were entered together into a final species distribution model, shown in Table 1.

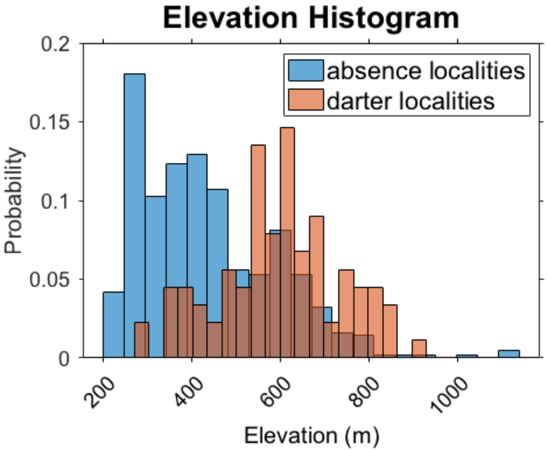
Category	**Variable name**	**Coefficient**	**Standard Error**	**P-value**
	Intercept	3.5305	1.2412	0.0045811
Topography				
	Elevation (m)	**4.5451**	**2.1814**	**0.037573**
	Slope (degrees)	**2.3938**	**0.88584**	**0.0**

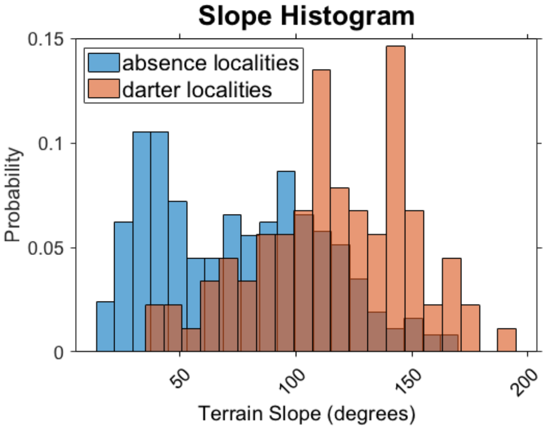
				070563
	Stream Gradient	0.02188	0.41505	0.95797
Geology				
	Geologic Age (Ma)	**3.6975**	**0.65046**	**1.94E-08**
	Soil Erosion (kg/ha/yr)	**1.1361**	**0.43461**	**0.0091398**
	Soil Water Content, July (%)	-4.2724	2.9795	0.15205
	Soil Water Content (%)	-2.8354	3.6246	0.43432
Rock type	Metasedimentary	-0.35393	3.3456	0.91578
	Metamorphic	3.1564	3.2883	0.33745
	Sedimentary Fine	-2.1515	1.3474	0.11077
	Sedimentary Dolomite	-2.117	1.499	0.15834
	Igneous Felsic	-7.8062	5.8873	0.1853
	Sedimentary Carbonate	-1.6006	1.592	0.31506
	Sedimentary Conglomerate	2.9241	2.4292	0.22911
Climate				
	Annual mean temperature (C)	**5.3717**	**2.126**	**0.011738**
	Annual Mean Precipitation (mm)	0.37044	1.4795	0.80237
	Precipitation, November	3.4589	1.8988	0.06894
	Actual Evapotranspiration (mm)	1.1721	4.5774	0.79798
	Potential Evapotranspiration (mm)	-4.681	4.8788	0.33767
	Upstream Snow Cover (%)	1.0688	0.56052	0.056954
Land use & Anthropogenic				
	Pasture Cover (%)	**3.7456**	**0.59815**	**6.70E-10**
	Upstream Pasture Cover (%)	**-2.8354**	**0.59851**	**2.63E-06**
	Gross Domestic Product	-1.3165	1.9249	0.49423
	Human Development Index	0.64111	1.801	0.72197

Hydrology				
	Annual Minimum Natural Discharge (m^3/s)	0.012976	0.39782	0.97399
	Land Surface Runoff (mm)	-0.28444	1.2978	0.82658
	Groundwater Table Depth (cm)	-0.16653	0.71626	0.81622

Table 1: Full model results for the final Species Distribution Model, sorted by variable category. Significant variables ($p < 0.05$) are shown in bold.

Annual mean temperature, elevation, pasture cover, geologic age, slope and soil erosion are positively correlated with occurrence of the Greenfin Darter. Upstream pasture cover is negatively correlated with the occurrence of the Greenfin Darter. Histograms for these variables, separated into presence localities and absence localities and normalized by probability, are shown in Figures 9a-9g.





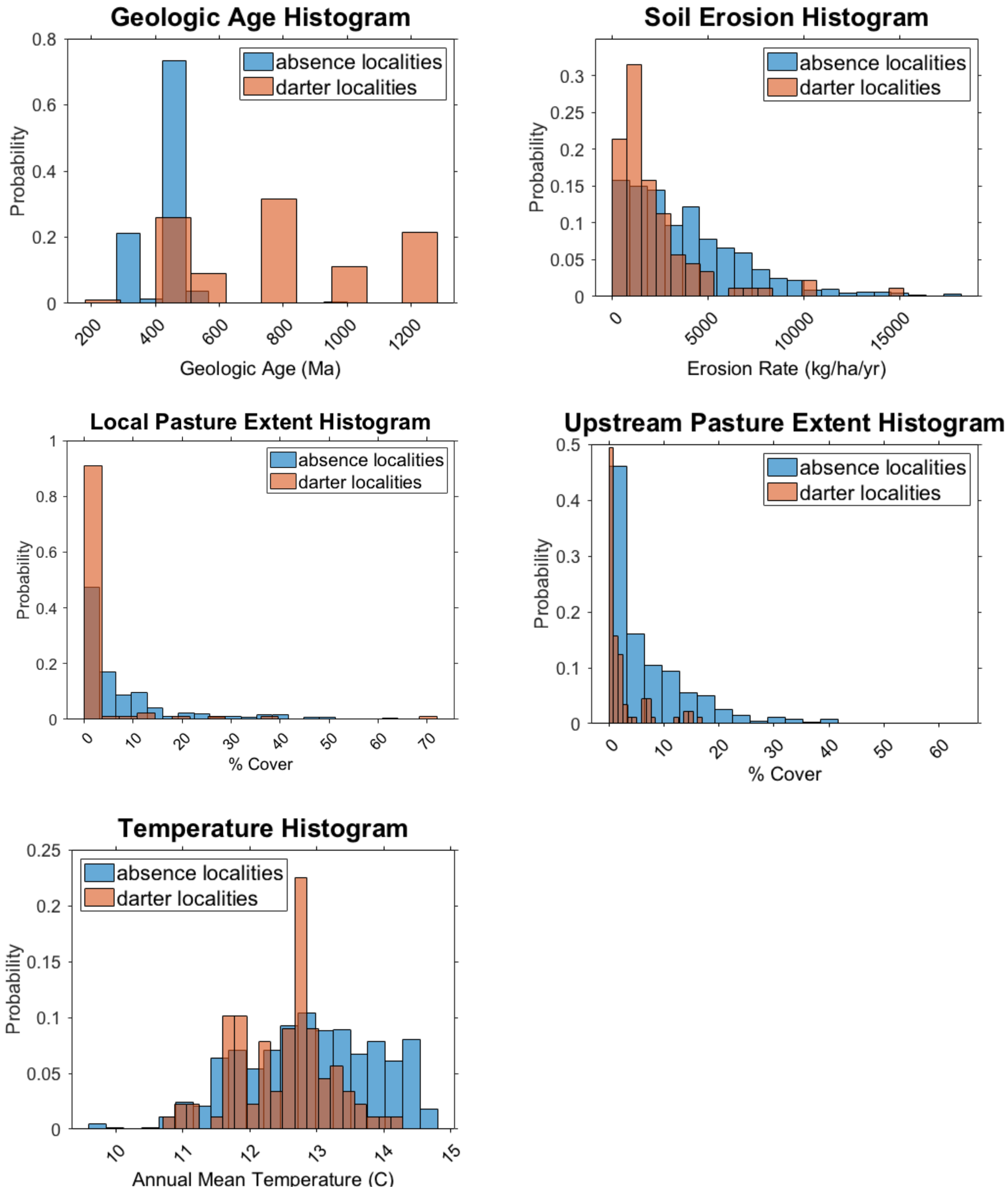


Figure 8: Histograms of final significant environmental variables separated by presence and absence points, normalized by probability. a) elevation (m), b) slope (degrees), c) geologic age (Ma), d) soil erosion (kg/ha/yr), e) local pasture cover (%), f) upstream pasture cover (%), g) temperature (C)

The R^2 value of this final model is 0.34. Thus, the model explains 34% of distribution of Greenfin Darter occurrences. This R^2 value is within range of similar studies, 0.3–0.8 (Rashleigh, 2004). A map of darter abundances predicted by the model is shown in Figure 9. Averaged over 1000 runs, the model correctly predicted 94.03% of Greenfin Darter presences, and 61.62% of Greenfin Darter absences.

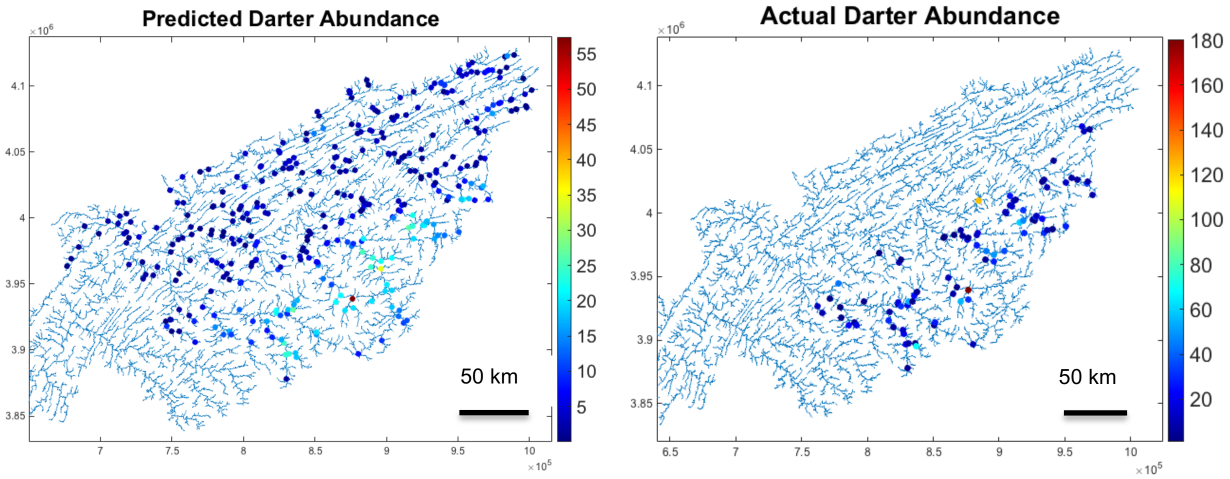


Figure 9a: Predicted Greenfin Darter abundances (≥ 1) across the Upper Tennessee River Basin, based on the final species distribution model. Figure 9b: Actual Greenfin Darter abundances (≥ 1).

Geology and topography

To further investigate Hypothesis 1, and to explore links between geological processes and biodiversity, I mapped knickpoint locations throughout the Upper Tennessee River Basin and examined their relations to Greenfin Darter populations. First, I found that 48% of darters live upstream of knickpoints, 28% live within knickpoints, and 24% live downstream of a knickpoint (Figure 10).

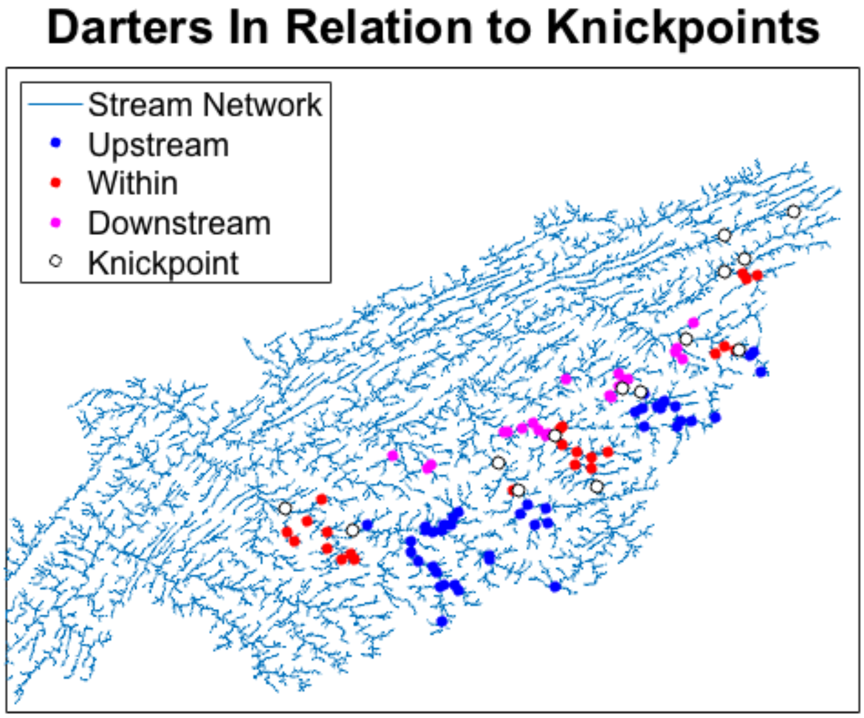


Figure 10: A map of knickpoints (shown in white), and darters in the Upper Tennessee River Basin. Blue points are locations upstream of a knickpoint where darters are found, red points are locations within a knickpoint, and magenta points are downstream of a knickpoint.

However, knickpoints throughout the basin are not identical. Thus, I compared the steepness of knickpoints in each tributary to the distribution of Greenfin Darters throughout that tributary. I found a positive correlation between the average stream gradient of each knickpoint against the proportion of darters living upstream of that knickpoint, with a correlation coefficient of 0.738 (Figure 11).

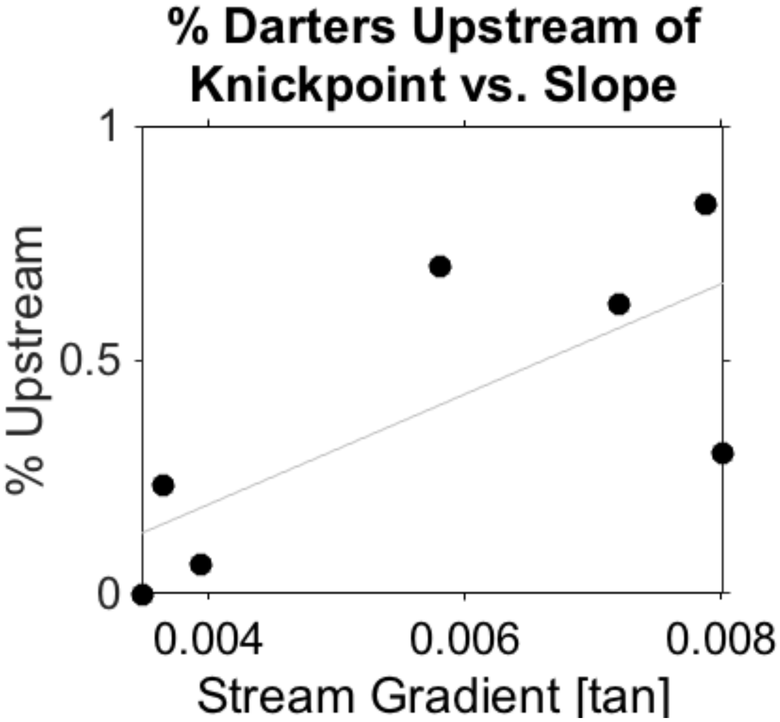


Figure 11: Each point corresponds to one of the tributaries in the Upper Tennessee Basin. There is a positive correlation between the slope of the knickpoint in each tributary and the percentage of darters living upstream of the knickpoint in that tributary.

Another way to examine the role of knickpoints in Greenfin Darter mobility, as well as in speciation processes, is by examining the genetic distances between individuals across a tributary. High genetic distance between pairs of individuals suggests that those two populations do not reproductively intermix. We have sufficient genetic data from the French Broad to compare genetic distances between pairs of individuals across the knickpoint in that tributary. Individuals located in the same ‘zone’ of the knickpoint (both within, both upstream or both downstream) were genetically closer than pairs of individuals in different zones of the knickpoint (Figure 12). There is a weak positive correlation between the streamwise distance and genetic distance between the individuals in each pair. However, the plot below shows two pairs of individuals (starred) that have similar streamwise distances separating them. Figure 12 shows that the pair in the same zone of the knickpoint are genetically closer than the pair in different zones of the knickpoint.

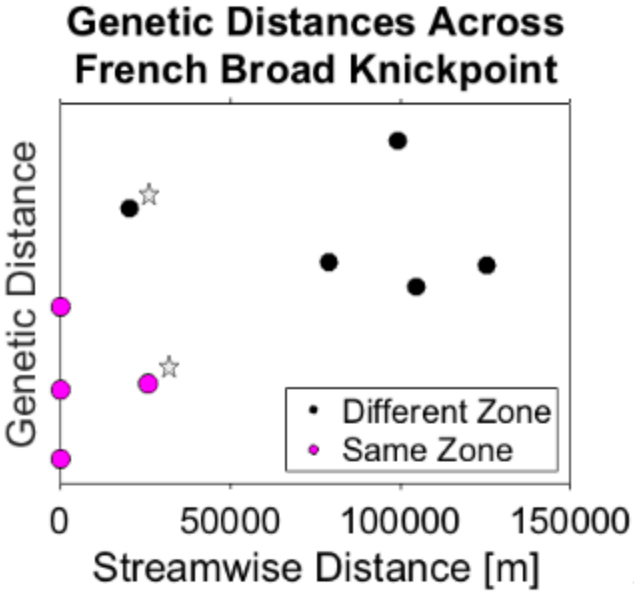


Figure 12: Each point represents a pair of individual Greenfin Darters from which genetic samples were collected. Magenta points represent pairs of darters on the same side of the knickpoint in their tributary (both within the knickpoint, both upstream or both downstream). Black points represent pairs of darters on different sides of the knickpoint (e.g. one upstream, one downstream). The x-axis represents streamwise distance between the pairs of darters in meters, and the y-axis is the genetic distance (dimensionless) between the genetic samples of the two individuals.

Discussion

Greenfin Darter distribution

1. *Topography*

My first hypothesis was that knickpoints constrain Greenfin Darter habitat by serving as barriers and isolating populations upstream. Thus, we would expect most individuals to live upstream of the knickpoints in their tributaries. However, 48% of darters live upstream of knickpoints, 28% live within knickpoints, and 24% live downstream of a knickpoint (Figure 10). Thus, knickpoints may not be absolute barriers to Greenfin Darter movement.

In fact, model results suggest that Greenfin Darters *prefer* to live in steeper sections of the river, since slope had a positive coefficient of 2.398 in the GLM (Table 2). This seems to contradict the hypothesis that knickpoints are too steep for darters to swim across. However, this result still suggests that topography is still important to understanding Greenfin Darter habitat, just not in the initially predicted direction. Greenfin Darters may live close to knickpoints not because knickpoints prevent them from migrating, but because the knickpoints themselves create a beneficial habitat for them. This is possible considering that darters have morphologically evolved to experience low drag in fast moving water (Carlson and Lauder, 2010). Thus, steep sections of the river could represent an ecological niche for the Greenfin Darter in the Tennessee River Basin.

On the other hand, as a plurality of Greenfin Darters live upstream of a knickpoint, they may still serve as a filter or partial barrier. If steep sections of the river are partial barriers to Greenfin Darter movement, we would expect tributaries with steeper knickpoints to have a higher proportion of individuals living upstream of the knickpoint than tributaries with flatter knickpoints. I found a positive correlation between the average stream gradient of each knickpoint against the proportion of Greenfin Darters living upstream of that knickpoint (Figure 11). This could suggest that the steepness of the river profile plays a role in Greenfin Darter mobility, and thus that knickpoints may serve as partial barriers in the range of the Greenfin Darter.

2. *Lithology*

My second hypothesis was that lithology constrains darter habitat; darters live where they do because they prefer to live on older, metamorphic rock. The positive coefficient of geologic age, which serves as a proxy for rock type, in the final model results supports this hypothesis. One possible explanation for why the Greenfin Darter could prefer metamorphic rock is that rock type can relate to differing degrees of channel-bed cover by sediment (Sklar and Dietrich, 2001), or varying concentrations of suspended sediment in the water (Kao & Milliman, 2008). Greenfin Darters deposit eggs in empty spaces within gravel riverbeds, and their eggs require "fresh, moving water" to properly grow and hatch; when fine sediment fills these empty spaces, fish eggs experience high mortality rates (Castro and Reckendorf, 1995). Thus, the presence of gravel bed load as well as concentrations of fine sediment could both impact the Greenfin Darter.

This explanation may initially seem to conflict with the positive coefficient for soil erosion in the model results, which would likely lead to more sediment in the water. However, the soil erosion variable does not distinguish between sizes and types of sediment being eroded. While increased amounts of fine sediment entering the river would negatively impact Greenfin Darter populations, increased bedload flux could have a positive impact. Additionally, it is possible that soil erosion has a positive coefficient because slope is correlated to erosion rate (Castro and Reckendorf, 1995). In the RiverATLAS dataset, the soil erosion variable comes from the Global Soil Erosion Modelling platform (GloSEM) v1.2, which utilizes the Universal Soil Loss Equation linking soil loss to rainfall erosivity, soil type, slope steepness, slope length, crop cover and soil management techniques (Borrelli et al, 2017).

Additionally, the impact of fine sediment on the Greenfin Darter aligns with the negative coefficient for upstream pasture cover, which correlates to higher runoff, erosion and sedimentation. The role of upstream and local pasture cover is discussed further in the Land Use section below. Ultimately, successful darter spawning requires a balance between discharge, velocity, and bed material (Castro and Reckendorf, 1995). Slope, lithology, soil erosion, and pasture cover all influence this balance.

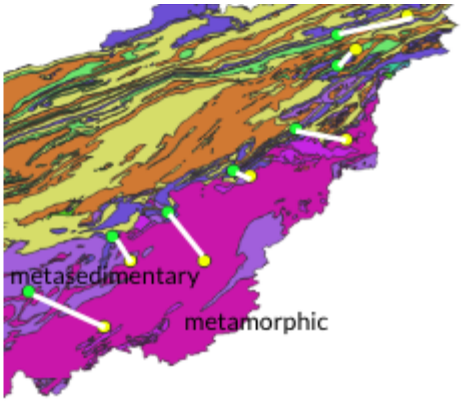


Figure 13: Knickpoint locations (white lines) overlaid over the geologic map of the Upper Tennessee River Basin.

The modern distribution data demonstrate that the Greenfin Darter lives in high-elevation, steep streams, primarily on old, metamorphic rock. However, many of these variables are spatially congruent. The boundary between rock types and the knickpoint locations are in approximately the same place. While the genetic data offers some promising support for the knickpoint hypothesis, the knickpoints themselves may be linked to differences in slope between different rock types. Thus, it is beyond the scope of this paper to make conclusions regarding whether darters prefer metamorphic rock, steep high-elevation areas, or whether they prefer some combination of these variables.

3. *Climate*

My third hypothesis was that the Greenfin Darter lives in higher-elevation areas because they prefer to live in a colder climate. However, temperature had a positive coefficient in the model results. Although this may initially seem contradictory to the fact that the Greenfin Darter lives in the higher-elevation and thus colder half of the upper Tennessee River Basin, the species seems to live in the warmer parts of that region (Figure 4b). Thus, one explanation is that climate influences where darters live *within* the upper half of the upper Tennessee River Basin. A study of heat tolerances of darter and minnow species in the southern Appalachians found that the Greenfin Darter had a higher warming tolerance than other species (Troia & Giam, 2019). While topography and geology seem to be the primary constraints on where darters live, climate could be a secondary driver influencing where darters live within these high-elevation, steep areas with underlying metamorphic rock.

4. *Land use*

Finally, my fourth hypothesis was that the Greenfin Darter range is controlled by anthropogenic land use. Upstream and local pasture cover both emerged as significant variables in the final species distribution model. The negative coefficient for upstream pasture cover supports the sedimentation explanation discussed above, as pasture cover causes increased sedimentation both directly and indirectly. As grazing animals search for water, they move towards rivers, destabilizing riverbanks as they trample vegetation (Castro and Reckendorf, 1995). Additionally, creating pasture typically requires clearing forest cover. A study of Southern Appalachian streams found that deforestation is linked to increased fine sediment content and decreased populations across the darter genus (Jones et al, 2001).

However, the model also had a positive coefficient for local pasture cover, which seems to conflict with the negative coefficient for upstream pasture cover. It is possible that local pasture cover could be correlated to other variables favorable to Greenfin Darter habitat, such as slope, since “increased runoff causes the hydrograph to become steeper” (Castro and Reckendorf, 1995). Additionally, histograms of these two variables show that Greenfin Darter populations are generally located in areas with low upstream *and* local pasture cover (Figure 8e-f). Regardless, the opposite coefficient signs of these closely related variables suggest that additional model calibration in the future could be useful.

Additionally, other anthropogenic variables in the model such as runoff, road density, population and urban development did not come up as significant in the final model. This differs from prior work on darters and similar species in the southeastern United States, which found anthropogenic land use to be the primary driver (Rashleigh, 2004). The role of land use and human activity is investigated further using additional data in the Anthropogenic Impacts section below.

Implications for evolution

The results of my species distribution and topographic analysis suggest that knickpoints may not relate to Greenfin Darter mobility in the way that I initially hypothesized, which has implications for the role of knickpoints in biodiversity. The presence of Greenfin Darters within and downstream of knickpoints in the Upper Tennessee River Basin, as well as the result that Greenfin Darters seem to prefer steeper habitats, does not support the hypothesis that knickpoints serve as geographic barriers causing allopatric speciation.

However, analysis of genetic data yields tentative evidence that knickpoints may still serve as barriers to gene flow. Figure 12, which plots genetic and streamwise distances between pairs of Greenfin Darters, shows two pairs of individuals (starred) that have similar streamwise distances separating them. The pair in the same zone of the knickpoint are genetically closer than the pair in different zones of the knickpoint. While there is not sufficient data to draw any significant conclusions, this initial pattern hints that despite the presence of individuals both upstream and downstream of the knickpoints in the Tennessee River Basin, gene flow across the knickpoint may be less than within the same zone of the knickpoint. Additionally, Figure 11 shows that more Greenfin Darters live upstream of knickpoints with steeper slopes. Although I only had sufficient genetic data to compare pairs of Greenfin Darters across the French Broad knickpoint, future work could test whether steeper knickpoints have a greater impact on gene flow than shallower knickpoints.

Finally, the abundance of endemic freshwater fish species in the Tennessee River Basin could be linked to an abundance of ecological niches. Rich variation in rock type (Figure 4d), geologic age, soil erosion and other variables (Figure 8) throughout the upper Tennessee River Basin could create varied habitat types for freshwater fish species. Freshwater fish species richness in the Tennessee River Basin is explored further in the following section.

Anthropogenic impacts

In this section, I explore relationships between human activity, Greenfin Darter occurrence, and overall freshwater fish biodiversity. First, I mapped dam locations and Greenfin Darter abundances in the Upper Tennessee River Basin. The Tennessee Valley Authority has constructed an extensive dam system along the Tennessee River for purposes including flood control, electricity generation, and job creation (Tennessee Valley Authority). However, dam construction often harms aquatic ecosystems by fragmenting habitat and increasing eutrophication, as well as harming marginalized communities. For example, the controversial construction of the Tellico Dam on the Tennessee River in 1979 caused the near-complete eradication of the endangered Snail Darter, and submerged seven historic Cherokee towns (Plater, 2013). Elsewhere in the United States, indigenous Yurok and Hoopa communities are currently fighting to remove dams along the Klamath River in California because of the dams' severe impacts on salmon populations, water quality and native livelihoods (Casarez, 2020).

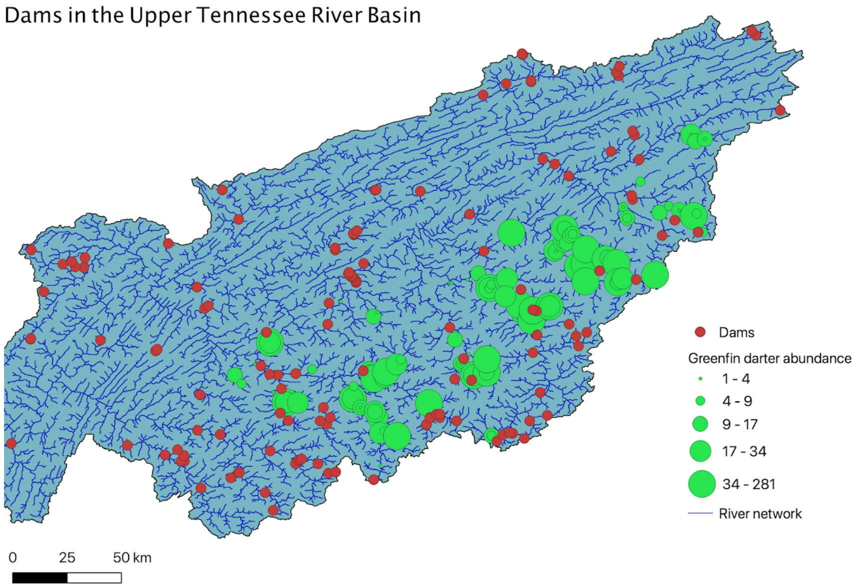
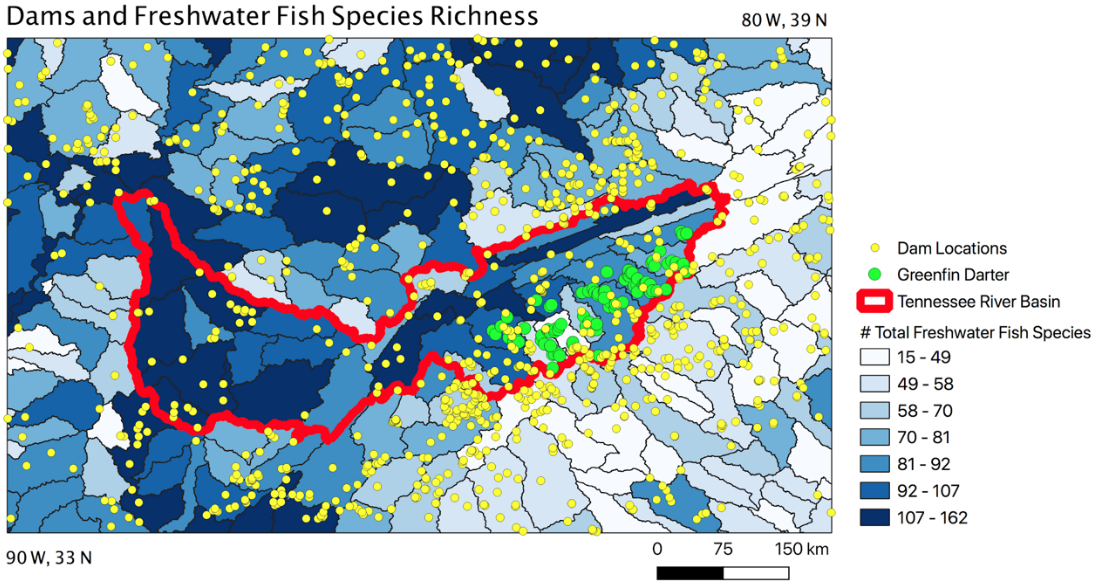


Figure 17: Dam locations (red) and Greenfin Darter locations (green), where point size corresponds to darter abundance

In the Upper Tennessee River Basin, dams seem to be spread out roughly evenly throughout the range of the Greenfin Darter (Figure 17). Thus, it is possible that dams have less impact on the Greenfin Darter. This aligns with species distribution modeling results, as the Degree of Regulation variable, a measure of how dams impact downstream river flow, did not emerge as significant during the modeling process (Appendix B). However, analysis of genetic data would provide more insight as to whether habitat fragmentation is occurring for the Greenfin Darter.

Next, to explore the impact of dams on biodiversity overall, I mapped the dam locations and freshwater fish species richness throughout the entire Tennessee River Basin and surrounding areas (Figure 18a). The correlation coefficient between the number of freshwater fish species and number of dams in a hydrologic unit was 0.0131, or no correlation. Additionally, I mapped dam locations and darter species richness specifically (Figure 18b). The correlation coefficient between the number of species in the darter genus and the number of dams in a hydrologic unit was 0.0102, also no correlation.



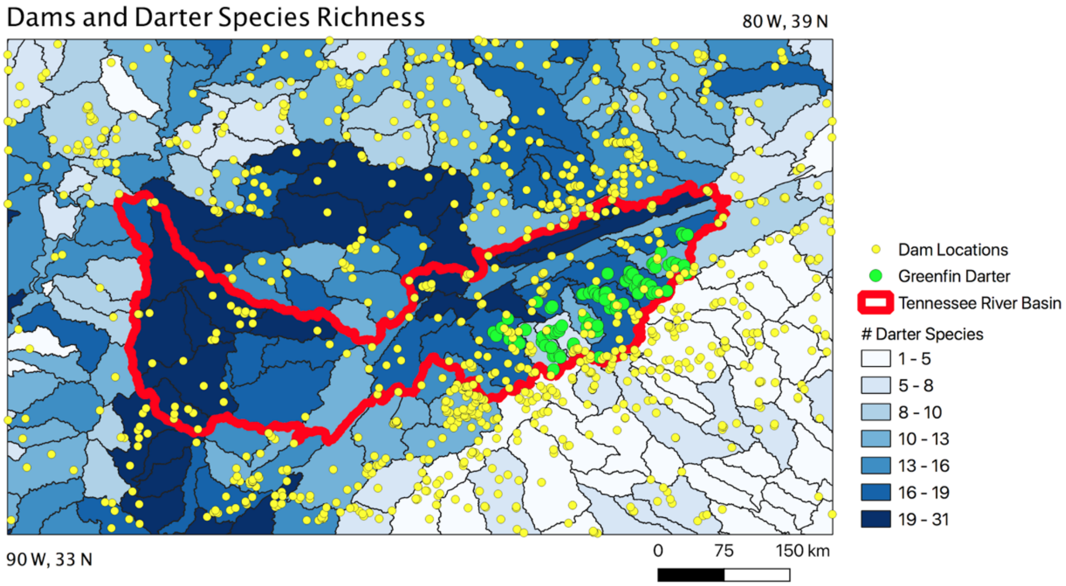


Figure 18: Dam locations and species richness in the southeastern United States.

Next, I examined the relationship between pollution, Greenfin Darter abundance and species richness. Model results and prior literature both indicate that darters are vulnerable to runoff, sediment and pollution. Point-source industrial and domestic pollution has specifically been found to impact fish assemblages in the upper Tennessee River Basin (Rashleigh, 2004). In the figure below, Superfund pollution site locations and their scores, which represent severity of contamination, are mapped with Greenfin Darter locations and species richness in the southeastern U.S. (Figure 19).

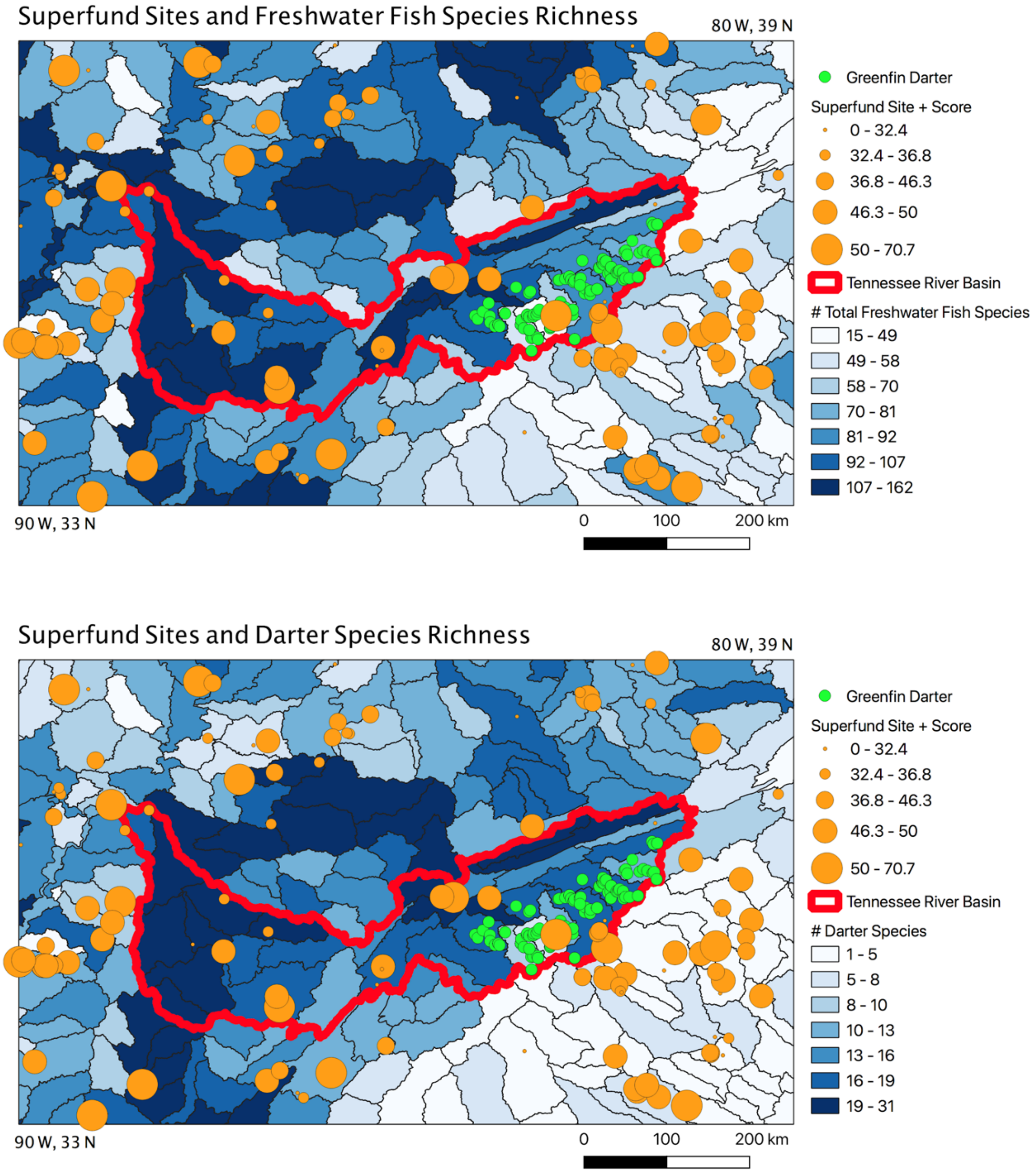


Figure 19: Superfund sites (orange), point size corresponding to site score, Greenfin Darter locations and overall freshwater fish species richness in the southeastern US.

The maps above show that the Upper Tennessee River Basin has 10 Superfund locations, 5 of which are upstream of where the Greenfin Darters reside (Figure 19). These locations are upstream of Greenfin Darter populations in the French Broad and Pigeon tributaries, aligning with prior literature that has found pollution impacts on darter assemblages in the French Broad tributary (Rashleigh, 2004). Furthermore, the Pigeon River is currently undergoing extensive cleanup and restoration, and fish populations are starting to rise (Pigeon River Recovery Project).

Overall, there is no correlation between the number of Superfund sites in a hydrologic unit and the overall species richness, with a correlation coefficient of -0.0231 (Figure 19a). However, Figure 19b shows that the hydrologic units with the highest number of darter species also generally have fewer Superfund sites than units with fewer species. Indeed, there is a weak negative correlation between the number of Superfund sites and the number of species in the darter genus, with a correlation coefficient of -0.1259 (Figure 19b). Thus, the presence of severe point pollution sources such as Superfund sites may negatively impact the darter genus as a whole. Since the Percidae (darters) family is one of the two most diverse fish groups in the Tennessee River Basin, and are proportionally the most imperiled, threats to darter species threaten the biodiversity of the region as a whole (Warren & Burr 1994).

The distribution of Superfund sites across the United States is not random. People of color are disproportionately impacted by environmental contamination; 49.8% of the people living within a mile of a Superfund site are people of color, and 28% of all people of color in the U.S. live within 3 miles of a Superfund site (U.S. EPA, 2020). This pattern is amplified in and around the Tennessee River Basin. In Tennessee, the only significant predictor of where Superfund sites are located is the percentage of Black people living in a census tract (McKane, 2016), and 55.9% of Black people in South Carolina live in a Superfund host census tract (Burwell-Naney et al, 2013). Longitudinal analysis of demographic composition around hazardous waste facilities suggests that the siting of hazardous waste facilities tends to target low-income communities of color, as opposed to the alternative explanation that demographic change occurs post-siting (Mohai & Saha, 2015). Furthermore, racism informs the Superfund cleanup process. Under the Trump administration, 61.1% of targeted sites were in majority white areas (Gibbs et al, 2019). Studies found that race and education level of surrounding neighborhoods impacted the speed at which Superfund sites were cleaned up at the beginning of the Superfund program, although this pattern may be beginning to shift (Burda and Harding, 2014).

In conclusion, decisions around where dams are built and where they are removed, as well as where contamination occurs and where it is cleaned up, impact both human and non-human communities in the Tennessee River Basin. Thus, biodiversity conservation and environmental justice are closely intertwined, and there are generative possibilities for collaboration and knowledge exchange between scientists, environmental justice advocates and impacted communities.

Conclusion

The Greenfin Darter prefers high-elevation, steep streams over older, metamorphic rock. Additionally, Greenfin Darters live in warmer regions of the upper Tennessee River Basin, and avoid areas downstream of pastures. Geology and topography, as well as climate and land use, all interact to shape sedimentation, water quality and dispersal pathways to create Greenfin Darter habitat in the Upper Tennessee River Basin.

These processes may also have an impact on biodiversity in the region generally. Analysis of genetic data across knickpoints suggests that topography may influence allopatric speciation in the Tennessee River Basin. Additionally, spatially variable lithology and corresponding sediment size and concentration across the basin may create varied ecological niches, further encouraging speciation. These results provide tentative evidence for links between geological processes and the generation of biodiversity in the Tennessee River Basin. On the other hand, the importance of temperature and pasture cover suggests that shifting climate conditions and land use patterns may reduce available habitats, threatening endemic species who have adapted to these specialized highland habitats.

This thesis suggests several future research directions. My genetic data was not extensive enough to draw significant conclusions about the role of knickpoints in genetic divergence; further genetic analysis could illuminate how knickpoints and other topographic features specifically impact gene flow at a fine-grained scale, and potentially lend insight into both the rich biodiversity of the Tennessee River Basin and the relationship between geology and speciation. Additionally, this analysis could be extended to include other species. I excluded the lower Tennessee River Basin in my analysis, because absence of the Greenfin Darter there may be due to competition with other species, rather than landscape features. However, extending the analysis to other darter species would allow one to explore how the landscape impacts not only the habitat and evolution of individual species, but ecological interactions as well. Furthermore, there are links between Superfund contamination sites and darter species richness, and between racial demographics and Superfund site locations. Future research should examine relationships between socioeconomic processes and biodiversity in the Tennessee River Basin, laying groundwork for ecology-environmental justice collaborations towards an environment where diverse human and non-human communities can thrive for generations to come.

Finally, environmental and earth science researchers have a responsibility to support the indigenous stewards of the places we study. The story of the co-evolution of the landscape and ecosystem in the Tennessee River Basin is incomplete without the history of stewardship by, displacement of and settler violence against the Eastern Band of Cherokee Indians (EBCI) and other indigenous groups. More information about the EBCI's fight for landback can be found here.

References

Albert, J., Craig, J., Tagliacollo, V., & Petry, P. (2017). *Upland and Lowland Fishes: A Test of the River Capture Hypothesis*.

Borrelli, P., Robinson, D. A., Fleischer, L. R., Lugato, E., Ballabio, C., Alewell, C., Meusburger, K., Modugno, S., Schütt, B., Ferro, V., Bagarello, V., Oost, K. V., Montanarella, L., & Panagos, P. (2017). An assessment of the global impact of 21st century land use change on soil erosion. *Nature Communications*, *8*(1), 2013. https://doi.org/10.1038/s41467-017-02142-7

Burda, M., & Harding, M. (2014). Environmental Justice: Evidence from Superfund cleanup durations. *Journal of Economic Behavior & Organization*, *107*, 380–401. https://doi.org/10.1016/j.jebo.2014.04.028

Burwell-Naney, K., Zhang, H., Samantapudi, A., Jiang, C., Dalemarre, L., Rice, L., Williams, E., & Wilson, S. (2013). Spatial disparity in the distribution of superfund sites in South Carolina: An ecological study. *Environmental Health*, *12*(1), 96. https://doi.org/10.1186/1476-069X-12-96

Carlson, R. L., & Lauder, G. V. (2009). Living on the bottom: Kinematics of benthic station-holding in darter fishes (Percidae: Etheostomatinae). *Journal of Morphology*, *271*(1), 25–35. https://doi.org/10.1002/jmor.10776

Casarez, I. (n.d.). *Local Tribes Sponsor Day of Action for Removal of Klamath Dams*. North Coast Journal. Retrieved November 2, 2020, from https://www.northcoastjournal.com/NewsBlog/archives/2020/10/22/local-tribes-sponsor-day-of-action-for-removal-of-klamath-dams

Castro, J., & Reckendorf, F. (1995). *Effects of Sediment on the Aquatic Environment* (Working Paper No. 6; Potential NRCS Actions to Improve Aquatic Habitat). National Resources Conservation Service. https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/technical/?cid=nrcs143_014201

Etheostoma chlorobranchium, Greenfin darter. (2012, February 1). Fishbase.Org. https://www.fishbase.in/summary/3407

Etnier, D., & Starnes, W. (1993). *The Fishes of Tennessee*. University of Tennessee Press. https://trace.tennessee.edu/utk_utpress/2

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Gallen, S. F. (2018). Lithologic controls on landscape dynamics and aquatic species evolution in post-orogenic mountains. *Earth and Planetary Science Letters*, *493*, 150–160. https://doi.org/10.1016/j.epsl.2018.04.029

Gibbs, L., Karlsson, M., & French, K. (2019). *Superfund: Polluters Pay so Children can Play*. Center for Health, Environment & Justice.

Hoorn, C., Perrigo, A., & Antonelli, A. (2018). *Mountains, climate and biodiversity*. Wiley Blackwell.

Horton, J. D. (2017). *The State Geologic Map Compilation (SGMC) Geodatabase of the Conterminous United States* [Data set]. U.S. Geological Survey. https://doi.org/10.5066/F7WH2N65

Jenkins, C. N., Van Houtan, K. S., Pimm, S. L., & Sexton, J. O. (2015). US protected lands mismatch biodiversity priorities. *Proceedings of the National Academy of Sciences*, *112*(16), 5081–5086. https://doi.org/10.1073/pnas.1418034112

Jones, E. B. D., Helfman, G. S., Harper, J. O., & Bolstad, P. V. (1999). Effects of Riparian Forest Removal on Fish Assemblages in Southern Appalachian Streams. *Conservation Biology*, *13*(6), 1454–1465. https://doi.org/10.1046/j.1523-1739.1999.98172.x

Kao, S., & Milliman, J. (2008). Water and Sediment Discharge from Small Mountainous Rivers, Taiwan: The Roles of Lithology, Episodic Events, and Human Activities. *Journal of Geology*, *116*(5), 431–448. https://doi.org/<p>10.1086/590921</p>

Kolmar, C. (2018, March 5). These Are The States With The Highest (And Lowest) Income Inequality. *Zippia*. https://www.zippia.com/advice/states-highest-lowest-income-inequality/

Lehner, B. (2019). *HydroATLAS Technical Documentation*. https://hydrosheds.org/images/inpages/HydroATLAS_TechDoc_v10.pdf

Lehner, B., Verdin, K., & Jarvis, A. (2008). New Global Hydrography Derived From Spaceborne Elevation Data. *Eos, Transactions American Geophysical Union*, *89*(10), 93–94. https://doi.org/10.1029/2008EO100001

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., & 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

Managing the Tennessee River for Multiple Benefits. (n.d.). Tennessee Valley Authority. Retrieved November 22, 2020, from https://www.tva.com/environment/managing-the-river

MATLAB (9.8.0 (R2020a)). (2020). [Computer software]. The MathWorks Inc. mathworks.com

McKane, R. G. (2016). *Toxic Waste & Race: The Changing Demographics Surrounding Superfund Sites in Tennessee* [Thesis]. https://ir.vanderbilt.edu/handle/1803/15183

Melo-Merino, S. M., Reyes-Bonilla, H., & Lira-Noriega, A. (2020). Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. *Ecological Modelling*, *415*, 108837. https://doi.org/10.1016/j.ecolmodel.2019.108837

Mohai, P., & Saha, R. (2015). Which came first, people or pollution? Assessing the disparate siting and post-siting demographic change hypotheses of environmental injustice. *Environmental Research Letters*, *10*(11), 115008. https://doi.org/10.1088/1748-9326/10/11/115008

NatureServe. 2010. Digital Distribution Maps of the Freshwater Fishes in the Conterminous United States. Version 3.0. Arlington, VA. U.S.A.

Pigeon River Recovery Project History. (2004). Pigeon River Recovery Project, University of Tennessee Knoxville. http://fwf.ag.utk.edu/Sites/pigeon/webpages/history.asp

Plater, Z. J. B. (2013). *The Snail Darter and the Dam: How Pork-Barrel Politics Endangered a Little Fish and Killed a River*. Yale University Press.

Population Surrounding 1,857 Superfund Remedial Sites. (2020). U.S. EPA, Office of Land and Emergency Management.

QGIS Geographic Information System (3.1.0). (2020). [Computer software]. QGIS Association. www.qgis.org

Rahbek, C., Borregaard, M. K., Colwell, R. K., Dalsgaard, B., Holt, B. G., Morueta-Holme, N., Nogues-Bravo, D., Whittaker, R. J., & Fjeldså, J. (2019). Humboldt's enigma: What causes global patterns of mountain biodiversity? *Science*, *365*(6458), 1108–1113. https://doi.org/10.1126/science.aax0149

Rashleigh, B. (2004). Relation of Environmental Characteristics to Fish Assemblages in the Upper French Broad River Basin, North Carolina. *Environmental Monitoring and Assessment*, *93*(1–3), 139–156.

Rawson, P. (2016). *Dams (National)*. Bureau of Transportation Statistics. https://catalog.data.gov/dataset/dams-national

Schwanghart, W., & Scherler, D. (2014). Short Communication: TopoToolbox 2 – MATLAB-based software for topographic analysis and modeling in Earth surface sciences. *Earth Surface Dynamics*, *2*(1), 1–7. https://doi.org/10.5194/esurf-2-1-2014

Scott, M. C. (2006). Winners and losers among stream fishes in relation to land use legacies and urban development in the southeastern US. *Biological Conservation*, *127*(3), 301–309. https://doi.org/10.1016/j.biocon.2005.07.020

Scott, M. C., & Helfman, G. S. (2001). Native Invasions, Homogenization, and the Mismeasure of Integrity of Fish Assemblages. *Fisheries*, *26*(11), 6–15. https://doi.org/10.1577/1548-8446(2001)026<0006:NIHATM>2.0.CO;2

Sklar, L. S., & Dietrich, W. E. (2001). Sediment and rock strength controls on river incision into bedrock. *Geology*, *29*(12), 1087–1090. https://doi.org/10.1130/0091-7613(2001)029<1087:SARSCO>2.0.CO;2

Stokes, M. F., Gallen, S. F., Sr., Perron, J. T., & Near, T. (2019). Knickpoints Drive Allopatric Speciation in the Blue Ridge Relict Landscape. *AGU Fall Meeting Abstracts*, *53*. http://adsabs.harvard.edu/abs/2019AGUFMEP53A..03S

Stokes, Maya F., & Perron, J. T. (2020). Modeling the Evolution of Aquatic Organisms in Dynamic River Basins. *Journal of Geophysical Research: Earth Surface*, *125*(9), e2020JF005652. https://doi.org/10.1029/2020JF005652

Szabo, I. (n.d.). *Greenfin Darter* [Photograph].

Troia, M. J., & Giam, X. (2019). Extreme heat events and the vulnerability of endemic montane fishes to climate change. *Ecography*, *42*(11), 1913–1925. https://doi.org/10.1111/ecog.04576

US EPA, O. (2015, September 4). *Search for Superfund Sites Where You Live* [Overviews and Factsheets]. US EPA. https://www.epa.gov/superfund/search-superfund-sites-where-you-live

US EPA, O. (2017, November 9). *What is Superfund?* [Overviews and Factsheets]. US EPA. https://www.epa.gov/superfund/what-superfund

Ward, A. D., & Trimble, S. W. (2003). *Environmental Hydrology, Second Edition*. CRC Press.

Warren, M., & Burr, B. (1994). Status of freshwater fishes of the United States: Overview of an imperiled fauna. *Fisheries*, *19*, 6–18.

Wobus, C., Whipple, K., Kirby, E., Snyder, N., Johnson, J., Spyropolou, K., Crosby, B., & Sheehan, D. (2006). Tectonics from topography: Procedures, promise and pitfalls. *Geological Society of America*, *398*, 55–74.

Appendix A: RiverATLAS environmental predictors

Adapted from the RiverATLAS v.10 Catalog (Lehner, 2019). References for source data can be found in Linke (2019).

Category	**Attribute**	**Source Data**	**Citation**	**Abbrev.**	**Variable Count**
Hydrology	Natural Discharge	WaterGAP v2.2	Doll et al. 2003	dis_m3	3
Hydrology	Land Surface Runoff	WaterGAp v2.2	Doll et al. 2003	run_mm	1
Hydrology	Inundation Extent	GIEMS-D15	Fluet-Chouinard et al. 2015	inu_pc	6
Hydrology	Liminicity (Percent Lake Area)	HydroLAKES	Messager et al. 2016	lka_pc	2
Hydrology	Lake Volume	HydroLAKES	Messager et al. 2016	lkv_mc	1
Hydrology	Reservoir Volume	GRanD v1.1	Lehner et al. 2011	rev_mc	1
Hydrology	Degree of Regulation	HydroSHEDS & GRanD	Lehner et al. 2011	dor_pc	1
Hydrology	River Area	HydroSHEDS & WaterGAP	Lehner & Grill 2013	ria_ha	2
Hydrology	River Volume	HydroSHEDS & WaterGAP	Lehner & Grill 2013	riv_tc	2
Hydrology	Groundwater Table Depth	Global Groundwater Map	Fan et al. 2013	gwt_cm	1
Physiography	Elevation	EarthEnv-DEM90	Robinson et al. 2014	ele_mt	4

Physiography	Terrain Slope	EarthEnv-DEM90	Robinson et al. 2014	slp_dg	2
Physiography	Stream Gradient	EarthEnv-DEM90	Robinson et al. 2014	sgr_dk	1
Climate	Climate Zones	GEnS	Metzger et al. 2013	clz_cl	1
Climate	Climate Strata	GEnS	Metzger et al. 2013	cls_cl	1
Climate	Air Temperature	WorldClim v1.4	Hijmans et al. 2005	tmp_dc	16
Climate	Precipitation	WorldClim v1.4	Hijmans et al. 2005	pre_mm	14
Climate	Potential Evapotranspiration	Global-PET	Zomer et al. 2008	pet_mm	14
Climate	Actual Evapotranspiration	Global Soil-Water Balance	Trabucco & Zomer 2010	aet_mm	14
Climate	Global Aridity Index	Global Aridity Index	Zomer et al. 2008	ari_ix	2
Climate	Climate Moisture Index	WorldClim & Global-PET	Hijmans et al. 2005	cmi_ix	14
Climate	Snow Cover Extent	MODIS/Aqua	Hall & Riggs 2016	snw_pc	15
Landcover	Land Cover Classes	GLC2000	Bartholome & Belward 2005	glc_cl	1
Landcover	Land Cover Extent	GLC2000	Bartholome & Belward 2005	glc_pc	44
Landcover	Potential Natural Vegetation Classes	EarthStat	Ramankutty & Foley 1999	pnv_cl	1
Landcover	Potential Natural Vegetation Extent	EarthStat	Ramankutty & Foley 1999	pnv_pc	30
Landcover	Wetland Classes	GLWD	Lehner & Doll 2004	wet_cl	1
Landcover	Wetland Extent	GLWD	Lehner & Doll 2004	wet_pc	22
Landcover	Forest Cover Extent	GLC2000	Bartholome & Belward 2005	for_pc	2
Landcover	Cropland Extent	EarthStat	Ramankutty & Foley 1999	crp_pc	2

Landcover	Pasture Extent	EarthStat	Ramankutty & Foley 1999	pst_pc	2
Landcover	Irrigated Area Extent (Equipped)	HID v1.0	Siebert et al. 2015	ire_pc	2
Landcover	Glacier Extent	GLIMS	GLIMS & NSIDC 2012	gla_pc	2
Landcover	Permafrost Extent	PZI	Gruber 2012	prm_pc	2
Landcover	Protected Area Extent	WDPA	IUCN & UNEP-WCMC 2014	pac_pc	2
Landcover	Terrestrial Biomes	TEOW	Dinerstein et al. 2017	tbi_cl	1
Landcover	Terrestrial Ecoregions	TEOW	Abell et al. 2008	tec_cl	1
Landcover	Freshwater Major Habitat Types	FEOW	Abell et al. 2008	fmh_cl	1
Landcover	Freshwater Ecoregions	FEOW	Hengl et al. 2014	fec_cl	1
Soils & Geology	Clay Fraction in Soil	SoilGrids1km	Hengl et al. 2014	cly_pc	2
Soils & Geology	Silt Fraction in Soil	SoilGrids1km	Hengl et al. 2014	slt_pc	2
Soils & Geology	Sand Fraction in Soil	SoilGrids1km	Hengl et al. 2014	snd_pc	2
Soils & Geology	Organic Carbon Content in Soil	SoilGrids1km	Hengl et al. 2014	soc_th	2
Soils & Geology	Soil Water Content	Global Soil-Water Balance	Trabucco & Zomer 2010	swc_pc	14
Soils & Geology	Lithological Classes	GLiM	Hartmann & Moosdorf 2012	lit_cl	1
Soils & Geology	Karst Area Extent	Rock Outcrops v3.0	Williams & Ford 2006	kar_pc	2
Soils & Geology	Soil Erosion	GloSEM v1.2	Borrelli et al. 2017	ero_kh	2
Anthropogenic	Population Count	GPW v4	CIESIN 2016	pop_ct	2
Anthropogenic	Population Density	GPW v4	CIESIN 2016	ppd_pk	2

Anthropogenic	Urban Extent	GHS S-MOD v1.0 (2016)	Pesaresi & Freire 2016	urb_pc	2
Anthropogenic	Nighttime Lights	Nighttime Lights v4	Doll 2008	nli_ix	2
Anthropogenic	Road Density	GRIP v4	Meijer et al. 2018	rdd_mk	2
Anthropogenic	Human Footprint	Human Footprint v2	Venter et al. 2016	hft_ix	4
Anthropogenic	Global Administrative Areas	GADM v2.0	University of Berkeley 2012	gad_id	1
Anthropogenic	Gross Domestic Product	GDP PPP v2	Kummu et al. 2018	gdp_ud	3
Anthropogenic	Human Development Index	HDI v2	Kummu et al. 2018	hdi_ix	1
Total	**56**				**281**

Appendix B: Full model results

Climate

Variable Name	Estimate	Standard Error	tStat	pValue
(Intercept)	2.3999	0.34846	6.8872	1.35E-11
clz_cl_cmj	-0.23718	1.0211	-0.23227	0.8164
cls_cl_cmj	1.2491	1.2834	0.97325	0.33079
tmp_dc_cyr	1.2398	2.1881	0.56661	0.57117
tmp_dc_uyr	-0.36361	3.0637	-0.11868	0.90556
tmp_dc_cmn	0	0	NaN	NaN
tmp_dc_cmx	0	0	NaN	NaN
tmp_dc_c01	-2.0426	8.4749	-0.24102	0.80961

tmp_dc_c02	14.352	9.6969	1.4801	0.13933
tmp_dc_c03	2.6314	9.3641	0.28101	0.77879
tmp_dc_c04	-9.8444	10.6	-0.92872	0.35338
tmp_dc_c05	0.93008	10.171	0.091445	0.92717
tmp_dc_c06	-19.361	11.088	-1.7461	0.081276
tmp_dc_c07	5.5578	10.057	0.55265	0.5807
tmp_dc_c08	-12.318	11.878	-1.037	0.30012
tmp_dc_c09	18.297	10.779	1.6974	0.090094
tmp_dc_c10	-4.8203	9.0439	-0.53299	0.59422
tmp_dc_c11	-0.53778	7.6839	-0.069988	0.94422
tmp_dc_c12	4.1488	6.2656	0.66216	0.50811
pre_mm_cyr	4.3864	3.0056	1.4594	0.14494
pre_mm_uyr	-7.4186	15.229	-0.48712	0.62634
pre_mm_c01	-14.181	10.021	-1.4151	0.15752
pre_mm_c02	3.2147	6.3386	0.50716	0.61222
pre_mm_c03	12.176	11.761	1.0353	0.30091
pre_mm_c04	0.012016	7.0618	0.0017015	0.99864
pre_mm_c05	-1.2511	6.2386	-0.20055	0.84111
pre_mm_c06	-3.6925	5.1472	-0.71739	0.4734
pre_mm_c07	4.9514	4.2756	1.1581	0.24726
pre_mm_c08	-4.0758	6.9846	-0.58355	0.55973
pre_mm_c09	5.4305	7.397	0.73415	0.46312
pre_mm_c10	-0.19201	5.6935	-0.033724	0.97311

pre_mm_c11	17.401	8.0947	2.1497	0.031952
pre_mm_c12	-1.7598	8.1545	-0.21581	0.82921
pet_mm_cyr	-10.333	8.4909	-1.2169	0.22408
pet_mm_uyr	9.859	6.457	1.5269	0.12728
aet_mm_cyr	-1.169	9.9435	-0.11757	0.90645
aet_mm_uyr	-10.463	5.6471	-1.8528	0.06436
ari_ix_cav	-18.946	10.561	-1.7939	0.073302
ari_ix_uav	23.376	12.562	1.8608	0.063225
cmi_ix_cyr	-6.9597	9.5876	-0.72591	0.46815
cmi_ix_uyr	0.70719	6.0333	0.11721	0.90673
cmi_ix_c01	1.7024	3.8457	0.44267	0.65815
cmi_ix_c02	7.7607	4.326	1.794	0.073284
cmi_ix_c03	-5.469	7.0555	-0.77514	0.43854
cmi_ix_c04	-1.1485	7.0627	-0.16262	0.87087
cmi_ix_c05	-1.032	6.9229	-0.14908	0.88154
cmi_ix_c06	4.1999	5.7957	0.72466	0.46892
cmi_ix_c07	-0.81169	5.0921	-0.1594	0.8734
cmi_ix_c08	2.7463	7.5776	0.36243	0.71715
cmi_ix_c09	-11.551	8.2604	-1.3984	0.16247
cmi_ix_c10	1.3218	7.2734	0.18174	0.85585
cmi_ix_c11	8.0529	5.9822	1.3461	0.17873
cmi_ix_c12	-8.677	5.0116	-1.7314	0.08386
snw_pc_cyr	-1.9219	1.5767	-1.2189	0.22332

snw_pc_uyr	-2.7257	1.0339	-2.6364	0.0085803
snw_pc_cmx	3.2346	2.3308	1.3878	0.16568
snw_pc_c01	-0.79315	1.3226	-0.59967	0.54894
snw_pc_c02	-1.1937	1.7068	-0.6994	0.48456
snw_pc_c03	0.012695	0.80765	0.015719	0.98746
snw_pc_c04	-0.24908	0.49775	-0.50042	0.61695
snw_pc_c05	-0.061629	0.42335	-0.14557	0.8843
snw_pc_c06	-0.94681	0.44648	-2.1206	0.034335
snw_pc_c07	0.37681	0.44823	0.84066	0.40085
snw_pc_c08	0.19049	0.39733	0.47943	0.6318
snw_pc_c09	-0.38577	0.4055	-0.95135	0.34178
snw_pc_c10	0.017001	0.51537	0.032987	0.97369
snw_pc_c11	-0.047875	0.6582	-0.072737	0.94204
snw_pc_c12	1.0728	1.0851	0.98861	0.32322

Hydrology

Variable Name	Estimate	Standard Error	tStat	pValue
(Intercept)	2.3999	0.39576	6.0641	2.18E-09
dis_m3_pyr	-23.287	47.53	-0.48994	0.62433
dis_m3_pmn	17.682	11.651	1.5177	0.12955
dis_m3_pmx	17.558	48.557	0.3616	0.71776
run_mm_cyr	3.1491	0.44902	7.0133	5.53E-12
inu_pc_cmn	-1.0396	3.3982	-0.30592	0.75976

inu_pc_umn	1.0893	2.7916	0.39023	0.69649
inu_pc_cmx	4.5966	5.2726	0.87181	0.38362
inu_pc_umx	1.4602	3.9979	0.36525	0.71504
inu_pc_clt	-3.691	4.5263	-0.81544	0.4151
inu_pc_ult	-3.1417	3.0469	-1.0311	0.30285
lka_pc_cse	-0.0018696	0.57017	-0.0032791	0.99738
lka_pc_use	-0.15464	0.53623	-0.28839	0.77313
lkv_mc_usu	3.7553	3.4807	1.0789	0.28102
dor_pc_pva	-1.6586	1.1636	-1.4254	0.15448
ria_ha_csu	-0.45202	1.1998	-0.37676	0.70647
ria_ha_usu	-0.74339	21.897	-0.033949	0.97293
riv_tc_csu	-0.20545	1.5544	-0.13217	0.89489
riv_tc_usu	-13.501	12.311	-1.0967	0.27317
gwt_cm_cav	1.6651	0.43917	3.7916	0.00016272

Landcover

Variable Name	Estimate	Standard Error	tStat	pValue
(Intercept)	2.3999	0.39023	6.1501	1.42E-09
glc_cl_cmj	-0.10184	0.91537	-0.11125	0.91145
glc_pc_c01	0	0	NaN	NaN
glc_pc_c02	-19.602	45.094	-0.43468	0.66395
glc_pc_c03	0	0	NaN	NaN
glc_pc_c04	-10.78	23.449	-0.45971	0.64589

glc_pc_c05	0	0	NaN	NaN
glc_pc_c06	-25.657	59.56	-0.43077	0.66679
glc_pc_c07	0	0	NaN	NaN
glc_pc_c08	0	0	NaN	NaN
glc_pc_c09	0	0	NaN	NaN
glc_pc_c10	0	0	NaN	NaN
glc_pc_c11	-1.4933	2.9458	-0.50694	0.61238
glc_pc_c12	0	0	NaN	NaN
glc_pc_c13	-5.2853	10.04	-0.52644	0.59878
glc_pc_c14	-0.044232	0.50811	-0.087052	0.93066
glc_pc_c15	0	0	NaN	NaN
glc_pc_c16	-33.696	64.186	-0.52498	0.59979
glc_pc_c17	0	0	NaN	NaN
glc_pc_c18	0	0	NaN	NaN
glc_pc_c19	0	0	NaN	NaN
glc_pc_c20	-3.8262	7.6304	-0.50144	0.61624
glc_pc_c21	0	0	NaN	NaN
glc_pc_c22	-13.813	26.5	-0.52126	0.60238
glc_pc_u01	-0.89289	0.66122	-1.3504	0.17741
glc_pc_u02	-13.006	23.961	-0.5428	0.58747
glc_pc_u03	0	0	NaN	NaN
glc_pc_u04	-4.9878	13.279	-0.37561	0.70734
glc_pc_u05	0	0	NaN	NaN

glc_pc_u06	-14.966	32.564	-0.45959	0.64598
glc_pc_u07	0	0	NaN	NaN
glc_pc_u08	0	0	NaN	NaN
glc_pc_u09	0	0	NaN	NaN
glc_pc_u10	0	0	NaN	NaN
glc_pc_u11	1.0062	2.1308	0.47223	0.63693
glc_pc_u12	0	0	NaN	NaN
glc_pc_u13	2.2398	2.8302	0.79137	0.42904
glc_pc_u14	0	0	NaN	NaN
glc_pc_u15	0	0	NaN	NaN
glc_pc_u16	32.847	38.714	0.84846	0.39652
glc_pc_u17	0	0	NaN	NaN
glc_pc_u18	0	0	NaN	NaN
glc_pc_u19	0	0	NaN	NaN
glc_pc_u20	1.2714	2.0262	0.62749	0.53058
glc_pc_u21	0	0	NaN	NaN
glc_pc_u22	14.703	17.39	0.84548	0.39818
pnv_cl_cmj	1.1747	1.6308	0.72033	0.4716
pnv_pc_c01	0	0	NaN	NaN
pnv_pc_c02	0	0	NaN	NaN
pnv_pc_c03	0	0	NaN	NaN
pnv_pc_c04	-0.010141	1.1796	-0.0085972	0.99314
pnv_pc_c05	-0.29974	1.8971	-0.158	0.87451

pnv_pc_c06	0	0	NaN	NaN
pnv_pc_c07	0	0	NaN	NaN
pnv_pc_c08	0	0	NaN	NaN
pnv_pc_c09	0	0	NaN	NaN
pnv_pc_c10	0	0	NaN	NaN
pnv_pc_c11	0	0	NaN	NaN
pnv_pc_c12	0	0	NaN	NaN
pnv_pc_c13	0	0	NaN	NaN
pnv_pc_c14	0	0	NaN	NaN
pnv_pc_c15	0	0	NaN	NaN
pnv_pc_u01	0	0	NaN	NaN
pnv_pc_u02	0	0	NaN	NaN
pnv_pc_u03	0	0	NaN	NaN
pnv_pc_u04	-15.054	41.757	-0.3605	0.7186
pnv_pc_u05	-37.294	109.82	-0.33958	0.73429
pnv_pc_u06	0	0	NaN	NaN
pnv_pc_u07	0	0	NaN	NaN
pnv_pc_u08	-38.442	109.54	-0.35093	0.72577
pnv_pc_u09	0	0	NaN	NaN
pnv_pc_u10	0	0	NaN	NaN
pnv_pc_u11	0	0	NaN	NaN
pnv_pc_u12	0	0	NaN	NaN
pnv_pc_u13	0	0	NaN	NaN

pnv_pc_u14	0	0	NaN	NaN
pnv_pc_u15	0	0	NaN	NaN
wet_pc_cg1	-20.159	51.558	-0.391	0.69594
wet_pc_ug1	3.3989	5.4434	0.62441	0.5326
wet_pc_cg2	0	0	NaN	NaN
wet_pc_ug2	0	0	NaN	NaN
wet_pc_c01	17.573	45.281	0.38809	0.69809
wet_pc_c02	0.47351	1.8977	0.24952	0.80304
wet_pc_c03	9.713	24.86	0.39071	0.69615
wet_pc_c04	0	0	NaN	NaN
wet_pc_c05	0	0	NaN	NaN
wet_pc_c06	0	0	NaN	NaN
wet_pc_c07	0	0	NaN	NaN
wet_pc_c08	0	0	NaN	NaN
wet_pc_c09	0	0	NaN	NaN
wet_pc_u01	-3.1	5.2544	-0.58997	0.55543
wet_pc_u02	0.016299	0.67672	0.024085	0.98079
wet_pc_u03	-0.35275	0.6025	-0.58547	0.55845
wet_pc_u04	0	0	NaN	NaN
wet_pc_u05	0	0	NaN	NaN
wet_pc_u06	0	0	NaN	NaN
wet_pc_u07	0	0	NaN	NaN
wet_pc_u08	0	0	NaN	NaN

wet_pc_u09	0	0	NaN	NaN
for_pc_cse	-13.404	81.032	-0.16542	0.86867
for_pc_use	49.075	46.514	1.0551	0.29182
crp_pc_cse	0.17664	0.8263	0.21377	0.8308
crp_pc_use	-1.0673	0.88962	-1.1998	0.23071
pst_pc_cse	3.6441	0.70951	5.1361	3.80E-07
pst_pc_use	-3.4268	0.71792	-4.7732	2.28E-06
ire_pc_cse	0.88634	0.52352	1.6931	0.090966
ire_pc_use	0.58216	0.52114	1.1171	0.26441
gla_pc_cse	0	0	NaN	NaN
gla_pc_use	0	0	NaN	NaN
prm_pc_cse	0	0	NaN	NaN
prm_pc_use	0	0	NaN	NaN
pac_pc_cse	-0.92873	0.63768	-1.4564	0.1458
pac_pc_use	1.1196	0.75197	1.4889	0.13704
tbi_cl_cmj	0	0	NaN	NaN
tec_cl_cmj	0.46199	0.55097	0.83849	0.40209
fmh_cl_cmj	0	0	NaN	NaN
fec_cl_cmj	0	0	NaN	NaN

Anthropogenic

Variable Name	Estimate	Standard Error	tStat	pValue
(Intercept)	2.3999	0.36763	6.528	1.29E-10

pop_ct_csu	-1.1938	1.9006	-0.62813	0.53012
pop_ct_usu	27.36	20.722	1.3204	0.18715
ppd_pk_cav	-1.095	1.1771	-0.93027	0.35255
ppd_pk_uav	0.99733	1.9363	0.51507	0.60667
urb_pc_cse	0.042608	1.1135	0.038265	0.96949
urb_pc_use	-0.17742	1.6062	-0.11046	0.91208
nli_ix_cav	-0.07203	1.906	-0.037791	0.96986
nli_ix_uav	-4.1796	2.1998	-1.9	0.05785
rdd_mk_cav	-0.085334	0.96188	-0.088716	0.92933
rdd_mk_uav	1.3831	1.7701	0.78138	0.43484
hft_ix_c93	4.9354	2.9488	1.6737	0.094641
hft_ix_u93	-1.5792	3.6309	-0.43494	0.66374
hft_ix_c09	-3.7562	3.3556	-1.1194	0.26336
hft_ix_u09	3.1951	3.9593	0.80698	0.41995
gad_id_cmj	0	0	NaN	NaN
gdp_ud_cav	-10.022	0.77921	-12.862	4.09E-34
gdp_ud_csu	0.9398	1.6982	0.55342	0.58015
gdp_ud_usu	-27.619	20.712	-1.3335	0.1828
hdi_ix_cav	9.4569	0.77808	12.154	5.89E-31

Physiography

Variable Name	Estimate	Standard Error	tStat	pValue
(Intercept)	2.3999	0.40145	5.9782	3.57E-09

elevation	2.3495	0.44515	5.278	1.74E-07
slp_dg_cav	0.84252	0.47175	1.7859	0.074535
sgr_dk_rav	-0.46191	0.43497	-1.0619	0.28863

Soils & Geology, including rock type

Variable Name	Estimate	Standard Error	tStat	pValue
(Intercept)	2.4522	1.271	1.9293	0.054105
cly_pc_cav	-1.2917	1.4317	-0.90221	0.36727
cly_pc_uav	2.7192	1.6413	1.6568	0.098023
slt_pc_cav	0.89337	1.1431	0.78151	0.43477
slt_pc_uav	1.4629	1.1075	1.321	0.18696
snd_pc_cav	-0.12579	1.8399	-0.068367	0.94551
snd_pc_uav	3.1035	2.001	1.551	0.12138
soc_th_cav	0.0062588	1.0104	0.0061945	0.99506
soc_th_uav	1.0393	1.3436	0.77347	0.43951
swc_pc_cyr	5.6669	4.4111	1.2847	0.19933
swc_pc_uyr	2.1694	1.2521	1.7327	0.083609
swc_pc_c01	-4.4553	2.69	-1.6563	0.098131
swc_pc_c02	-0.088343	0.73353	-0.12044	0.90417
swc_pc_c03	0	0	NaN	NaN
swc_pc_c04	1.2322	1.2546	0.98208	0.32641
swc_pc_c05	-3.2935	2.517	-1.3085	0.19116
swc_pc_c06	1.3095	3.685	0.35536	0.72243

swc_pc_c07	-7.7211	3.9059	-1.9768	0.048472
swc_pc_c08	5.288	4.2882	1.2331	0.21795
swc_pc_c09	0.9306	5.0028	0.18601	0.85249
swc_pc_c10	-4.1642	5.2906	-0.78709	0.4315
swc_pc_c11	-4.3772	5.3676	-0.81548	0.41508
swc_pc_c12	7.0489	4.3303	1.6278	0.10404
lit_cl_cmj	0.48362	0.45695	1.0584	0.29026
kar_pc_cse	0.80845	0.65352	1.2371	0.21649
kar_pc_use	-0.55455	0.71628	-0.7742	0.43908
ero_kh_cav	1.9144	0.55266	3.464	0.00056571
ero_kh_uav	-1.0894	0.5831	-1.8682	0.06216
geo_age_ma	3.6782	0.64671	5.6875	1.92E-08
RockUnits_metasedimentary	-0.69926	2.9472	-0.23726	0.81253
RockUnits_metamorphic	2.491	3.2076	0.77661	0.43766
RockUnits_RockUnits__sedimentary_fine	-0.67402	1.3897	-0.485	0.62783
RockUnits_sedimentary_dolomite	-0.37873	1.604	-0.23612	0.81341
RockUnits_igneous_felsic	-6.8803	6.1801	-1.1133	0.26597
RockUnits_sedimentary_carbonate	0.079437	1.695	0.046867	0.96263
RockUnits_sedimentary_conglomerate	2.4557	2.5856	0.94976	0.34257