

Comparative and connected weight of evidence analyses to increase confidence in directions for
Topeka Shiner conservation.

by

Shelby Kuck

B.S., Samford University, 2023

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Division of Biology
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Approved by:

Major Professor
Dr. Martha Mather

Copyright

© Shelby Kuck 2026.

Abstract

Biodiversity monitoring data provides an important source of ecological evidence about native and invasive species trends. A high priority for many conservation agencies is to understand how habitat use, biotic interactions, and human impacts affect declining and uncommon fish, especially those listed as threatened or endangered species. However, quantitative analysis of biodiversity datasets at the state-wide or regional scales that are useful for management is a challenge. More specifically high variability, small sample sizes, uncontrolled underlying heterogeneity, and the violation of statistical assumptions can increase ambiguous results between studies, mask important trends, create spurious statistical patterns, or increase unassigned error. These quantitative difficulties can prevent conservation professionals from learning all they can about existing data. *The purpose of this research is to modify and test a systematic approach to analyzing existing field monitoring data to extract insights, identify gaps, and guide future actions for research and management.* For this purpose, we use the specific test case of Topeka Shiner (*Notropis topeka*), a native Great Plains fish species that is listed as federally-endangered and Kansas-threatened. Chapter 1 illustrates how comparing outputs of multiple statistical methodologies applied to the same dataset (multiple logistic regression vs random forest analyses) can corroborate consistent ecological trends on which professionals can build and diagnose inconsistent results that require further examination. Chapter 2 applies an existing 12-step adaptive management framework to Topeka Shiner data to advance understanding and guide future conservation actions. This iterative framework uses weight-of-evidence to integrate existing literature, visualizations, statistical analyses, and ecological interpretation, and guide future data and restoration action for a focused overarching question, taxa and scale. As a result of applying this framework to a rare taxon, useful insights

into Topeka Shiner distributions emerged for ecological trends with previously described (*stream depth*, *species richness*, *number of dams*), and new (clay soil content, or *clay*) environmental variables. This research tested additional emerging questions through iterative application of the above-described framework. Comparing the outputs from models created to test these new data-driven questions underscored the importance of *predator species richness*, nonlinear (quadratic) relationships with *baseflow index*, and interactions between *clay* and *baseflow index* across river basins. By building upon the original framework in successive iterations, we identified useful and testable questions regarding the abiotic and biotic variables that impact Topeka Shiner presence. Information generated from our analyses can guide future data actions, collaborative agency review processes for fish protections (e.g., KDWP, EPA, USFWS, NMFS), and subsequent framework cycles. This research offers examples of practical pathways for researchers to leverage limited data and maximize accrual of knowledge for fish ecology and conservation.

Table of Contents

List of Figures	viii
List of Tables	xv
Acknowledgements	xxiii
Dedication	xxv
Preface	xxvi
Chapter 1 - Advancing ecological insights to inform conservation management of the threatened Topeka Shiner: Integrating multiple statistical approaches applied to long-term monitoring data in KS	1
Abstract	1
Introduction	3
Methods	6
Study Organism	6
Fish Data	7
Environmental Data	8
Multiple Logistic Regression (<i>Objective 1a</i>)	8
MLR Variable Importance	1
MLR Response Trends	2
Random Forest (<i>Objective 1b</i>)	2
RF Variable Importance	3
RF Response Trends	3
Comparison and Integration Across Statistical Approaches (<i>Objective 2</i>)	4
Variable Importance	4
Response Trends	4
Map Predictions	4
Results	5
Multiple Logistic Regression (<i>Objective 1a</i>)	5
Random Forest (<i>Objective 1b</i>)	6
Comparisons (<i>Objective 2</i>)	7
Variable Importance	7

Response Trends	8
Spatial Trends	8
Discussion.....	9
Agreement in Methodological Approaches Can Build Confidence (<i>Objectives 1, 2, 3</i>)	9
Disagreement in Statistical Outcomes for Ecological Drivers of Distribution Guides Future Work (<i>Objectives 1, 2, 3</i>)	11
Integrating Approaches (<i>Objectives 2, 3</i>)	14
Conclusions.....	16
Acknowledgments	17
Tables.....	19
Figures	24
Chapter 2 - Applying and expanding a structured approach to the systematic statistical analysis and ecological interpretation of existing data to guide imperiled Great Plains fish conservation	32
Abstract.....	32
Introduction.....	34
Importance of Data-Based Evidence for Effective Natural Resource Conservation.	34
Challenges with Analyzing Existing Monitoring Data.	34
Great Plains Stream Fish.....	35
Objectives.....	37
Methods	38
Overview of the 12 Steps in the Rode et al. (2025) Framework.....	38
Applying the Framework to a Rare Taxon (Objective 1).	39
Framework Iterations (Objective 2; Figure 1).	42
Compare Data Visualizations Systematically and Critically (Objective 2A).	42
Exploring Nonlinear Variable Parameterization (Objective 2B).	43
Refine Specific Metrics within an Ecological variable (Objective 2C).	43
Explore Interactions (Objective 2D).	44
Account for Random Effects (Objective 2E).	45
Compare Models (Objective 2F).	46
Results.....	46

Objective 1 – Application of the Framework to a New Taxon (Steps, 3, 7, 8A, 8C, 8E-F, 9).	46
Objective 2A. Compare Data Visualizations Systematically and Critically.....	49
Objective 2B. Explore Nonlinear Variable Parameterization.	54
Objective 2C. Refine Metrics within an Ecological variable.....	55
Objective 2D. Explore Interactions.....	56
Objective 2E. Account for Random Effects.....	56
Objective 2F. Compare Models.	57
Discussion.....	59
Overview.....	59
Take-Home 1: Applicability of the Original Framework to Topeka Shiner Data (Objective 1).	59
Take-Home 2: Multiple Connected Framework Iterations can Enhance Ecological Insights (Objective 2).	61
Take-Home 3: Benefits For Research and Management– Objective 3, Framework Steps 10-12.....	65
Conclusions - Advantages of Our Process.....	70
Acknowledgements.....	71
Tables.....	72
Figures	85
References.....	98
Appendix A - Chapter 1	118
Tables.....	118
Appendix B - Chapter 2	120
Tables.....	120
Figures	134

List of Figures

- Figure 1.1. Distribution of Topeka Shiner presences in stream assessment samples among (A) major river basins in (B) Kansas, USA..... 24
- Figure 1.2. Marginal effects on the predicted probability of Topeka Shiner presence (Y axis) from the top multiple logistic regression model that investigates 11 ecological variables. Predictor ecological variables on the X axis include: (A) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (B) *stream depth* (estimated in meters), (C) *clay* (% in catchment surface geology), (D) *number of dams* (total in the mainstem), (E) *species richness* (number of fish species in a single sample), (F) *gradient* (m/m), (G) *grassland* (% cover in catchment), (H) *cropland* (% cover in the catchment), (I) *curvature*, (J) *development* (% cover in the catchment), and (K) *pastureland* (% cover in the catchment). Raw data distributions are depicted with black points at the top and bottom of each panel, response trends are black lines, and confidence intervals are shown in grey. The model results in this figure correspond to Table 2..... 25
- Figure 1.3. Importance of 11 ecological variables describing Topeka Shiner population distributions in Kansas from 670 stream samples collected between 1947 and 2020 using (A) mean decrease in accuracy and (B) mean decrease in Gini purity index modeled by random forest on the X axes. Predictor ecological variables on the Y axes include: *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *stream depth* (estimated in meters), *clay* (% in catchment surface geology), *number of dams* (total in the mainstem), *species richness* (number of fish species in a single sample), *gradient* (m/m), *grassland* (% cover in catchment), *cropland* (% cover in the catchment), *curvature*, *development* (% cover in the catchment), and *pastureland* (% cover in the catchment)..... 26
- Figure 1.4. Partial Dependency of predicted probability of Topeka Shiner presence from 11 ecological variables. Predictor ecological variables include: (A) *species richness* (number of fish species in a single sample), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *curvature*, (E) *development* (% cover in the catchment), (F) *gradient* (m/m), (G) *number of dams* (total in the mainstem), (H) *grassland* (% cover in catchment), (I) *stream*

depth (estimated in meters), (J) *cropland* (% cover in the catchment), and (K) *pastureland* (% cover in the catchment). Raw data distributions are shown as black points, raw response trends as black lines, and LOESS smoothed trends with confidence intervals in grey. (B). 27

Figure 1.5. Multiple logistic regression and random forest model performance as described by (A), (C) confusion matrices with predicted values on the X axis and observed values on the Y axis; and (B), (D) receiving operator characteristic (ROC) curves with 1-specificity on the X axis, sensitivity on the Y axis, model predictions in solid, and expected values in dashed lines. Multiple logistic regression accuracy = 88.21%. Area under the curve (AUC) = 0.935. Random forest out-of-bag error = 7.16%. Accuracy = 92.84%. AUC = 0.998. 28

Figure 1.6. Predicted probability of presence in response to clay substrate composition (%) and species richness across the state of KS from (A), (C) multiple logistic regression and (B), (D) random forest ecological models of Topeka Shiner distributions in Kansas from 670 stream samples collected between 1947 and 2020. Predictor ecological variables include: (A), (B) *clay* (% in catchment surface geology) and (C), (D) *species richness* (number of fish species in a single sample) on the X axes. Response on the Y axes represents predicted probability of Topeka Shiner presence. 29

Figure 1.7. Predicted probability of presence in response to stream depth (m), number of dams, and baseflow index across the state of KS from (A), (C), (E) multiple logistic regression and (B), (D), (F) random forest ecological models of Topeka Shiner distributions in Kansas from 670 stream samples collected between 1947 and 2020. Predictor ecological variables include: (A), (B) *stream depth* (estimated in meters); (C), (D) *number of dams* (total in the mainstem), and (E), (F) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow) on the X axes. Response on the Y axes represents predicted probability of Topeka Shiner presence. 30

Figure 1.8. Map of predicted probability of Topeka Shiner presence in Kansas from 12,086 stream samples collected (81% between 2000-2021) as modeled by 11 ecological variables using (A) multiple logistic regression and (B) random forest. Comparisons were made between statistical approaches for sites above and below 50% probability of presence from multiple logistic regression and predicted presence (1) or absence (0) from random forest. 31

Figure 2.1. Modified Rode et al. (2025) adaptive management framework showing the objectives addressed here. In our first objective, we apply the original framework to the Topeka Shiner

(*Notropis topeka*), using a statewide Kansas monitoring database. The 12 steps are described in the text. In the second objective, we test six data-driven questions in multiple iterations of the 12-step cycle. These connected iterations include (A) comparing data to identify agreement and disagreement in trends, (B) testing nonlinear relationships, (C) refining ecological variable metrics, (D) testing interactions between ecological variables, (E) incorporating random effects, and (F) comparing statistical models to identify consistently influential variables. For our third objective, we integrate each iteration of the 12-step framework to develop ecological insights and recommendations for future data collection and conservation actions (discussion). 85

Figure 2.2. *Objective 1 – Step 3*. Number of instances in Topeka Shiner literature review of 35 articles where the ecological categories of (A) system size, (B) substrate, (C) flow, (D) biota, (E) land use, and (F) dams were discussed or investigated in Kansas and other states. Percent of statistically significant ($p \leq 0.05$; +/-) and nonsignificant (NS; including qualitative or untested) trends included below with letters (*a- ah) corresponding to citations in Appendix Table 2. 86

Figure 2.3. *Objective 1 – Step 8A*. Map of the major river basins in Kansas (numbered 1-12) with points to represent stream biodiversity sample sites ($N=12,086$) collected between 1857 and 2021. Eighty-one percent of data were collected between 2000-2021. Topeka Shiner was present at sites labeled with red points ($n=341$), while the fish was absent among sites in blue. 87

Figure 2.4. *Objective 1 – Step 8C*. Kernel Density Estimate (KDE) ridge plots of eleven ecological variables (X axis) calculated from the original data. Variables shown include (A) *stream depth* (estimated in meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). The Y Axis is the probability density of a Topeka Shiner presence ($n = 335$; red) or absence (blue) sample occurring at a given value of ecological variables. Data were collected from 670 stream sites in Kansas from 1947 to 2020. The area under each curve is equal to 1. 88

Figure 2.5. *Objective 2A – Step 9*. Multi-panel plot of *stream depth* (X axis) to evaluate consistency of trends in the Step 8 quantitative sub-steps (8A, 8B, 8F). The Y axis on panels represent (A) histogram frequencies (Sub-step 8B), (B) boxplots of Topeka Shiner presence ($n = 335$; 1) and absence (0) (Sub-step 8C), (C) KDE probability of fish presence (red) and absence (blue) (Sub-step 8C), (D) distributional habitat maps of Kansas, and predicted probability of Topeka Shiner presence from (E) single (Sub-step 8C), and (F) multiple logistic regression (Sub-step 8F). Data were collected from 670 stream sites in Kansas from 1947 to 2020..... 89

Figure 2.6. *Objective 2A – Step 9*. Multi-panel plot of *baseflow index* (X axis) to evaluate consistency of trends in the Step 8 quantitative sub-steps (8A, 8B, 8F). The Y axis on panels represent (A) histogram frequencies (Sub-step 8B), (B) boxplots of Topeka Shiner presence ($n = 335$; 1) and absence (0) (Sub-step 8C), (C) KDE probability of fish presence (red) and absence (blue) (Sub-step 8C), (D) distributional habitat maps of Kansas, and predicted probability of Topeka Shiner presence from (E) single (Sub-step 8C), and (F) multiple logistic regression (Sub-step 8F). Data were collected from 670 stream sites in Kansas from 1947 to 2020..... 90

Figure 2.7. *Objective 2B – Step 8D*. Predicted probability of Topeka Shiner presence (Y axis) from single logistic regression models assessing the role of (A) a single linear parameter, (B) a single squared parameter, and (C) two quadratic parameters for the *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow) ecological variable (X axis). All three model estimations significantly ($p \leq 0.05$) describe Topeka Shiner presence ($n = 335$); however, the quadratic parameterization was most parsimonious ($\Delta AIC_c < 2$) and best describes the data pattern. Data were collected from 670 stream sites in Kansas from 1947 to 2020. 91

Figure 2.8. *Objective 2B – Step 8F*. Marginal effects on the predicted probability of Topeka Shiner presence (Y axis) from the top multiple logistic regression model that investigates the 11 ecological variables defined in Objective 1 plus *baseflow index*². Predictor ecological variables include: (A) *stream depth* (estimated in meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development*

(% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). Raw data distributions are depicted with black points at the top and bottom of each panel, response trends are black lines, and confidence intervals are shown in grey. The model results in this figure correspond to Table 3. Data were collected from 670 stream sites in Kansas from 1947 to 2020..... 92

Figure 2.9. *Objective 2C – Step 8B-D*. Multi-panel plots with histogram sample frequencies (A-D), KDE probability ridge plots (E-H) with Topeka Shiner presence ($n = 335$; red) and absence (blue), boxplots of fish presence (1) and absence (0) (I-L), and predicted probability of fish presence from single logistic regression (M-P) on the Y axis. The four biotic variables on the X axes include *species richness* (number of fish species in a single sample), *predator species richness* (predators), *competitor species richness* (competitors), and *facilitator species richness* (facilitators) for communities sampled at 670 stream sites in Kansas between 1947 and 2020. 93

Figure 2.10. *Objective 2C – Step 8F*. Marginal effects on the predicted probability of Topeka Shiner presence (Y axis) from the top multiple logistic regression model for biotic investigation. Predictor ecological variables include: (A) *stream depth* (estimated in meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). Raw data distributions are depicted with black points at the top and bottom of each panel, response trends are black lines, and confidence intervals are shown in grey. The model results in this figure correspond to Table 2. Data were collected from 670 stream sites in Kansas from 1947 to 2020..... 94

Figure 2.11. *Objective 2D – Step 8F*. Marginal effects on the predicted probability of Topeka Shiner presence (Y axis) from the interaction *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow) and *clay* (% in catchment surface geology). Predictor variables in the model include: (A) *stream depth* (estimated in meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of

groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). The model results in this figure correspond to Table 3. X axes show (A) *clay* and (B) *baseflow index*. Legends correspond to the multiple response curves as (A) *baseflow index* and (B) *clay* are held at specific levels. Data were collected from 670 stream sites in Kansas from 1947 to 2020. 95

Figure 2.12. *Objective 2E – Step 8B*. (A) Map of major river basins (Basin) in Kansas (numbered 1-12). (B) Percentage (Y axis) of 670 stream sample sites sampled between 1947 to 2020 with Topeka Shiner presence ($n = 335$; red) and absence (blue) among the twelve major river basins (X axis). 96

Figure 2.13. *Objective 2E – Step 8F*. Modeled predicted Topeka Shiner probability of presence (0 to 1; blue to red points) from mixed multiple logistic regression estimates (Objective 2E) for 12,086 stream sites in Kansas, USA, collected between 1857 and 2021. Eighty-one percent of data were collected between 2000-2021. Map axes correspond to latitude and longitude. The model results in this figure correspond to Table 4. 97

Appendix B Figure 1. *Objective 1 – Step 8B*. Histogram sample frequencies (Y axis) of eleven ecological variables associated with 670 stream site samples in Kansas from 1947 to 2020, including (A) *stream depth* (estimated in meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem)..... 134

Appendix B Figure 2. *Objective 1 – Step 8D*. Boxplots of eleven ecological variables (X axis) including (A) *stream depth* (estimated in meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development*

(% cover in the catchment), (I) <i>cropland</i> (% cover in the catchment), (J) <i>pastureland</i> (% cover in the catchment), and (K) <i>number of dams</i> (total in the mainstem). Data distributions are split by factor of Topeka Shiner presence (1) and absence (0) on the Y axis from collections in Kansas between 1947 and 2020.....	135
Appendix B Figure 3. <i>Objective 1 – Step 8D</i> . Single logistic regression response trends for the impact of eleven ecological variables including (A) <i>stream depth</i> (estimated in meters), (B) <i>clay</i> (% in catchment surface geology), (C) <i>baseflow index</i> (% contribution of groundwater and other sustained water sources to stream flow), (D) <i>species richness</i> (number of fish species in a single sample), (E) <i>grassland</i> (% cover in catchment), (F) <i>gradient</i> (m/m), (G) <i>curvature</i> , (H) <i>development</i> (% cover in the catchment), (I) <i>cropland</i> (% cover in the catchment), (J) <i>pastureland</i> (% cover in the catchment), and (K) <i>number of dams</i> (total in the mainstem). The response is the predicted probability of Topeka Shiner presence (Y axis) in Kansas from 1947 to 2020.	136
Appendix B Figure 4. <i>Objective 2E – Step 8F</i> . Intercept adjustments (X axis) for the random effect of twelve major river basins (Y axis) in KS to the modeled predicted probability of Topeka Shiner presence from 1947 to 2020.	137

List of Tables

Table 1.1. Eleven ecological variables used to model Topeka Shiner distributions in Kansas. Included are variable name, relevance to fish, definition of measurement, units, spatial scale, database used, and source.	19
Table 1.2. Results from multiple logistic regression on standardized data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas from 1947 to 2020 ($n=670$). Included are variable name, coefficient (β), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p -value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor variables include <i>clay</i> (% in catchment surface geology), <i>stream depth</i> (estimated in meters), <i>number of dams</i> (total in the mainstem), <i>species richness</i> (number of fish species in a single sample), <i>development</i> (% cover in the catchment), <i>grassland</i> (% cover in catchment), <i>pastureland</i> (% cover in the catchment), <i>baseflow index</i> (% contribution of groundwater and other sustained water sources to stream flow), <i>curvature</i> , <i>cropland</i> (% cover in the catchment), and <i>gradient</i> (m/m).....	22
Table 1.3. Comparison of importance (ranks 1-11 from most to least important) for 11 ecological variables as identified by significance and the absolute magnitude of multiple logistic regression (MLR) estimates and mean decrease in accuracy of random forest (RF) models. Significant MLR estimates ($p \leq .05$) are denoted with *.	23
Table 2.1. Variables within 8 ecological categories used to model Topeka Shiner distributions in Kansas. Included are variable name, relevance to fish, definition of measurement, units, spatial scale, database used, and source.	72
Table 2.2. Conceptual progression of framework iterations examining Topeka Shiner distributions in Kansas by research objective, model iteration (Steps 7-9), future directions (Steps 10-11), and applications (Steps 11-12). Included are summaries of research questions, motivations, predictions, results, emerging questions, proposed future measurements, priority locations, research importance, and management implications associated with each objective. New questions addressed by existing data have “not applicable” (N/A) listed in remaining sections. Examples of future directions and applications provided are non-exhaustive.	75

Table 2.3. Objective 1 – Step 8F. Results from multiple logistic regression on standardized data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas ($n=335$) from 670 stream site samples collected between 1947 and 2020. Included are variable name, coefficient (β), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor variables include *stream depth* (estimated in meters), *clay* (% in catchment surface geology), *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *species richness* (number of fish species in a single sample), *grassland* (% cover in catchment), *gradient* (m/m), *curvature*, (H) *development* (% cover in the catchment), *cropland* (% cover in the catchment), *pastureland* (% cover in the catchment), and *number of dams* (total in the mainstem). These calculations correspond to Step 8F of the Rode et al. 2025 framework. . 79

Table 2.4. *Objective 2B – Step 8F*. Results from multiple logistic regression on standardized data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas ($n=335$) from 670 stream site samples collected between 1947 and 2020, with an additional parameter added to describe a quadratic relationship with *baseflow index* (underlined). Included are variable name, coefficient (β), standard error (SE), Z-statistic (Z), lower and upper 95% confidence interval, *p*-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor variables include *stream depth* (estimated in meters), *clay* (% in catchment surface geology), *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *species richness* (number of fish species in a single sample), *grassland* (% cover in catchment), *gradient* (m/m), *curvature*, *development* (% cover in the catchment), *cropland* (% cover in the catchment), *pastureland* (% cover in the catchment), and *number of dams* (total in the mainstem). All other methodological details and variables are the same as described in Objective 1. The results in this table correspond to Fig. 8. These calculations correspond to Step 8F of the Rode et al 2025 framework. 80

Table 2.5. *Objective 2C – Step 8F*. Results from multiple logistic regression on standardized data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas ($n=335$) from 670 stream site samples collected between 1947 and 2020, with predator species richness (underlined). Included are variable name,

coefficient (β), standard error (SE), Z-statistic (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor variables include *stream depth* (estimated in meters), *clay* (% in catchment surface geology), *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *predator species richness* (number of fish species in a single sample which are potential predators of Topeka Shiner), *grassland* (% cover in catchment), *gradient* (m/m), *curvature*, *development* (% cover in the catchment), *cropland* (% cover in the catchment), *pastureland* (% cover in the catchment), and *number of dams* (total in the mainstem). All other methodological details and variables are the same as described in Objective 1. The results in this table correspond to Fig. 10. These calculations correspond to Step 8F of the Rode et al. 2025 framework. 81

Table 2.6. *Objective 2D – Step 8F*. Results from multiple logistic regression on standardized data using 5 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas ($n=335$) from 670 stream site samples collected between 1947 and 2020, with an interaction term included between *baseflow index* and *clay*. Eighty one percent of the data was collected between 2000-2021. Included are variable name, coefficient (β), standard error (SE), Z-statistic (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor variables include *stream depth* (estimated in meters), *clay* (% in catchment surface geology), *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *predator species richness* (number of fish species in a single sample which are potential predators of Topeka Shiner), and *number of dams* (total in the mainstem). All two and three way interactions were tested but only the interaction shown below is significant. The reduced model excludes noncontributing variables. All other methodological details and variables are the same as described in Objective 1. The new variable metric added is underlined. The results in this table correspond to Fig. 11. These calculations correspond to Step 8F of the Rode et al. 2025 framework. 82

Table 2.7. *Objective 2E – Step 8F*. Results from multiple logistic regression on standardized data using 5 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas ($n=670$), with a random effect of major river basins. Eighty one percent of the data was collected between 2000-2021. Included are variable name,

coefficient (β), standard error (SE), Z-statistic (Z), lower and upper 95% confidence interval, p -value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF) for fixed effects and standard deviation (SD) and lower and upper 95% confidence interval for random intercepts. Predictor variables include *stream depth* (estimated in meters), *clay* (% in catchment surface geology), *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *predator species richness* (number of fish species in a single sample which are potential predators of Topeka Shiner), and *number of dams* (total in the mainstem). All other methodological details and variables are the same as described in Objective 2D. The results in this table correspond to Fig. 13.

These calculations correspond to Step 8F of the Rode et al. 2025 framework. 83

Table 2.8. <i>Objective 2F – Step 8F</i> . Comparison of standardized estimate magnitudes from candidate multiple logistic regression models predicting the probability of Topeka Shiner presence in Kansas ($n=335$) from 670 stream site samples collected between 1947 and 2020. Included are top models for Objectives 1-2E, corrected Akaike Information Criterion (AIC_c), variable names, beta coefficients (β), statistical significance (asterisk to denote $p \leq 0.05$), and relative ranks of significant variables within models (1 = highest magnitude of influence on predicted Topeka Shiner presence). Variables shown include <i>stream depth</i> (estimated in meters), <i>clay</i> (% in catchment surface geology), <i>baseflow index</i> (% contribution of groundwater and other sustained water sources to stream flow), <i>species richness</i> (number of fish species in a single sample), <i>predator species richness</i> (number of fish species in a single sample which are potential predators to Topeka Shiner), <i>grassland</i> (% cover in catchment), <i>gradient</i> (m/m), <i>curvature, development</i> (% cover in the catchment), <i>cropland</i> (% cover in the catchment), <i>pastureland</i> (% cover in the catchment), and <i>number of dams</i> (total in the mainstem). Blanks (--) for terms not included in models. All β are relative to the individual model intercept.	84
---	----

Appendix A Table 1. Estimates from multiple logistic regression on original data using 11 ecological explanatory variables to model the response of Topeka Shiner (<i>Notropis topeka</i>) probability of presence in Kansas from 1947 to 2020 ($n=670$). Significant estimates ($p \leq 0.05$) are denoted with an *.	118
---	-----

Appendix A Table 2. Summary statistics (mean $\pm$ standard deviation (SD)) for variables included in models of Topeka Shiner (<i>Notropis topeka</i>) probability of presence in Kansas from 1947 to 2020 ($n=670$).....	119
Appendix B Table 1. List of fish species recorded in the Kansas Stream Assessment and Monitoring dataset used in the calculation of species richness in analyses. Included are common name, scientific name, taxonomic family, and assigned trait guild (if applicable). Species are listed with trait guilds of potential competitors (with invertivore diets), predators (large piscivores), and facilitators (nest building and egg guarding) of Topeka Shiner as assigned by the FishTraits Database and expert advice. Species that are not assigned to a guild are left blank (--).	120
Appendix B Table 2. <i>Objective 1 – Step 3</i> . Summary of ecological variables investigated in studies of Topeka Shiner in Kansas and other states arranged by year and author names. Included are the study system, type of paper contribution, and ecological variables investigated if relevant. The direction of statistically significant trends in (+/-). Variables shown include <i>stream size</i> , <i>substrate</i> , <i>groundwater</i> and <i>stream flow</i> , <i>biota</i> , <i>gradient</i> , <i>development agriculture</i> , and <i>dams</i> . Web of Science keywords “Topeka Shiner” (“ <i>Notropis topeka</i> ”) and “Kansas.” The alphabetical labels correspond with Fig. 2. Articles with ecological variables beyond the scope of our study are left blank (--).	126
Appendix B Table 3. <i>Objective 1 – Step 8</i> . Results from multiple logistic regression on original data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas ($n=670$). Eighty-one percent of the data was collected between 2000-2021. Included are variable name, coefficient (b), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor ecological variables include: (A) <i>stream order</i> (modified Strahler), (B) <i>clay</i> (% in catchment surface geology), (C) <i>baseflow index</i> (% contribution of groundwater and other sustained water sources to stream flow), (D) <i>species richness</i> (number of fish species in a single sample), (E) <i>grassland</i> (% cover in catchment), (F) <i>gradient</i> (m/m), (G) <i>curvature</i> , (H) <i>development</i> (% cover in the catchment), (I) <i>cropland</i> (% cover in the catchment), (J) <i>pastureland</i> (% cover in the catchment), and (K) <i>number of dams</i> (total in the mainstem). These calculations correspond	

to Step 8F of the Rode et al. 2025 framework. $\Delta AIC_c = 3.22$. The new variable metric added is underlined. 128

Appendix B Table 4. *Objective 1 – Step 8*. Results from multiple logistic regression on original data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas (n=670). Eighty-one percent of the data was collected between 2000-2021. Included are variable name, coefficient (b), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor ecological variables include: (A) *stream velocity* (average annual gage estimate in feet per second), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). These calculations correspond to Step 8F of the Rode et al. 2025 framework. $\Delta AIC_c = 6.24$. The new variable metric added is underlined. 129

Appendix B Table 5. *Objective 2B – Step 8*. Results from multiple logistic regression on original data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas (n=670). Eighty-one percent of the data was collected between 2000-2021. Included are variable name, coefficient (b), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor ecological variables include: (A) *stream depth* (meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). These calculations correspond to Step 8F of the Rode et al. 2025 framework. The new variable metric added is underlined. 130

Appendix B Table 6. *Objective 2C – Step 8*. Results from multiple logistic regression on original data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas (n=670). Eighty-one percent of the data was collected between 2000-2021. Included are variable name, coefficient (b), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor ecological variables include: (A) *stream depth* (meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *predator species richness* (number of fish species in a single sample which are potential predators of Topeka Shiner), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). These calculations correspond to Step 8F of the Rode et al. 2025 framework. The new variable metric added is underlined. 131

Appendix B Table 7. *Objective 2D – Step 8*. Results from multiple logistic regression on original data using 5 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas (n=670). Eighty-one percent of the data was collected between 2000-2021. Included are variable name, coefficient (b), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor ecological variables include: (A) *stream depth* (meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), and (E) *number of dams* (total in the mainstem). These calculations correspond to Step 8F of the Rode et al. 2025 framework. The new interaction between *clay* and *baseflow index* added is underlined. .. 132

Appendix B Table 8. *Objective 2E – Step 8*. Results from multiple logistic regression on original data using 5 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas (n=670). Eighty-one percent of the data was collected between 2000-2021. Included are variable name, coefficient (b), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF) for fixed effects and standard deviation

(SD) and 95% CI for random intercepts. Predictor ecological variables include: (A) *stream depth* (meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), and (E) *number of dams* (total in the mainstem). These calculations correspond to Step 8F of the Rode et al. 2025 framework. 133

Acknowledgements

I would first like to extend my thanks to Dr. Martha Mather who has always desired to make me a better fisheries professional. She has often encouraged me to squeeze the most out of every opportunity and challenge in my research.

This project was funded through the Kansas Department of Wildlife and Parks. The project was administered by the Kansas Cooperative Fish and Wildlife Research Unit as a joint effort between Kansas State University, the U.S. Geological Survey, U.S. Fish and Wildlife Service, the Kansas Department of Wildlife and Parks (KDWP), and the Wildlife Management Institute. I am fortunate to train as a future management and research professional under the care of these collaborators. I would like to thank the Kansas State Department of Biology for accepting me into the program and providing access to the many resources afforded to its students.

I would like to thank the Kansas Department of Wildlife and Parks for entrusting me with their fish biodiversity monitoring dataset. I offer many thanks to Ryan Waters, KDWP stream biologist, and the rest of the Stream Assessment team for collecting the survey data utilized here. I am grateful for the opportunity to work with this data which took a tremendous amount of time and effort to amass. A special thanks to Jordan Hofmeier and the Ecosystem Services Division for always supporting our progress on the project and answering my many questions about agency work. I would additionally like to acknowledge the contributions and immense support of Mark Van Scoyoc and Trevor Starks, providing helpful information about restoration and monitoring efforts and feedback on deliverables.

Thank you to each of my excellent committee members. Dr. Trisha Moore – I greatly enjoyed working together to address data limitations and explore Kansas subsurface hydrology.

To Dr. Kate Nelson: it was an honor to participate in your last cohort of students in the geospatial analysis course at K-State; thank you for helping me translate my skills from ArcGIS to R programming and sharing your lab's expertise and code for our data processing pipeline. Thank you to Dr. Devon Oliver, for your statistical and fish management expertise— I greatly appreciate your guidance and collaboration.

Thanks also to Dr. Dana Infante and Dr. Jared Ross from Michigan State University for providing access to datasets. Suyash Pratap: thank you for offering your valuable time to the project – I enjoyed getting to collaborate with you. Finally, immense thanks to Jason Luginbill and Gibran Suleiman for sharing your expertise and lending an ear to discuss Topeka Shiner populations. I hold you both in the highest regards.

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government

Dedication

To He who sustains me.

Preface

The contents of this thesis represent ideas and approaches developed in collaboration with my major professor, members of my thesis committee, and state agency partners. All data presented are resultant of collections performed by the Kansas Department of Wildlife and Parks Stream Assessment and Monitoring Program. Although the work is my own, the chapters are presented in the third person. Appropriate keywords for this document include stream fish ecology, endangered species act, habitat restoration, species distribution models, and ecological monitoring. Chapter 1 has been prepared for general audiences including fisheries professionals and could be submitted to *Fisheries Magazine* or another fish-specific journal (*Transactions of the American Fisheries Society*, *Fisheries Management and Ecology*). Chapter 2 is formatted for a more applied conservation audience and could be submitted to the *North American Journal of Fisheries Management* or other agency-facing journals (*Fisheries Management and Ecology*, *Aquatic Conservation: Marine and Freshwater Ecosystems*).

Chapter 1 - Advancing ecological insights to inform conservation management of the threatened Topeka Shiner: Integrating multiple statistical approaches applied to long-term monitoring data in KS

Abstract

Conservation of native species at statewide and regional scales is advanced by reliable ecological insights. Yet not all research studies can provide easily generalizable evidence for habitat associations and threats for the state and regional scales at which conservation decisions are made by agencies. Research studies that use statewide monitoring data, on the other hand, can assess a range of trends at larger spatial scales, which is often consistent with the needs of state or regional conservation. However, the characteristics of statewide monitoring datasets conflict with the common assumptions of traditional statistical analyses, presenting challenges for developing clear ecological interpretations of trends. Here, we ask how analysts can extract more useful ecological insights from large field datasets despite unavoidable statistical challenges. Our target taxon was Topeka Shiner (*Notropis topeka*, $n = 335$ presences), a federally endangered and Kansas-threatened native minnow. For the purpose of identifying consistent and divergent ecological trends and guiding future research and management for this species, we compared and integrated insights from complementary analyses of Kansas statewide monitoring data. We applied both multiple logistic regression and random forest models to the same monitoring dataset and 11 environmental and biotic variables. The output of both statistical analyses agreed on the importance of two ecological variables (*clay* and *species richness*), while our results for three other variables (*stream depth*, *number of dams*, and *baseflow index*) diverged. Our process for integrating results across analyses provided a useful first step in

distinguishing consistent and dissimilar trends. In summary, our approach advanced our ecological understanding of Topeka Shiner distributions from statewide monitoring data by strengthening our confidence in ecological insights and suggesting new directions for research and restoration based on both similar and divergent trends.

Introduction

Conservation of threatened and endangered species advances when professionals acquire clear and actionable, ecological insights (Cook et al., 2013). Ecological understanding of where and why rare organisms occur can provide a solid foundation for evidence-based decision-making by conservation professionals. For conservation professionals working with declining taxa, data on habitat conditions and species distributional patterns are especially useful at larger spatial scales corresponding to state and federal management jurisdictions (Brooks et al., 2004; Guisan et al., 2013). For state agencies, wide coverage of data at large spatial scales can support the assessment of species trends across regions. Survey data from state-wide monitoring programs can provide information on landscape-scale patterns and trends in species presence-absence, abundance, and diversity (Lindenmayer et al., 2022). While collecting data using field surveys is essential for supporting conservation management and decision-making, some challenges in using monitoring data can hinder progress toward achieving these goals.

When applied to field data, current statistical approaches can result in unsatisfying analytical outcomes for resource managers for several reasons. First, statistical assumptions are often not met by even the best monitoring datasets (Flather, 1997; Gitzen et al., 2012), and as a result, model predictions for accuracy and precision can provide erroneous field expectations for species presence or detection. Second, the spatial scale at which field data are collected can affect the interpretability and utility of field data (Olsen et al., 1997). Third, statewide monitoring may not always capture detailed ecological patterns that are important for understanding species distributions, especially when trading data resolution for greater coverage or targeting rare species (Nielsen et al., 2009; Potter et al., 2016). Fourth, the unavoidable rarity of threatened and endangered species exacerbates many statistical issues due to small sample sizes (Olden &

Jackson, 2002). Finally, analyses of monitoring data can yield ambiguous, erroneous, or conflicting ecological insights without adequate consideration of the original study design and analytical intent (Hutto & Belote, 2013), especially when applied *post-hoc* (Gitzen et al., 2012). Finer-scale, more-controlled experimental field data collection can systematically eliminate some concerns about assumption violations, but smaller or single-system studies may not provide the landscape-scale perspectives needed for state and regional conservation decisions (Englund & Cooper, 2003; Ommer & Perry, 2022). These characteristic problems encountered when analyzing monitoring datasets raise the overarching question we address: how can analysts extract more useful ecological insights from large field datasets despite these statistical challenges?

Here, we utilize an approach for enhancing the use of statewide biodiversity monitoring data that leverages multi-method analyses to support *weight of evidence* integration of ecological insights (Mather et al., 2026). Specifically, we identify consistent and divergent ecological trends by comparing fundamentally different statistical analyses that are applied to the same dataset and conservation questions. The Mather et al. (2026) approach identifies new ecological insights by employing both multiple logistic regression (MLR) and random forest (RF) analyses. Considered a traditional regression modeling approach, MLR can identify influential habitat variables by providing direct and more interpretable estimates of ecological relationships (Hosmer et al., 2013). However, MLR makes many assumptions (e.g. measurement without error, linearity, absence of outliers) about data that are unrealistic for field surveys (Bolker, 2008). Meanwhile RF algorithms, like other machine learning approaches, can generate more robust predictions by repeatedly classifying data and reporting averaged trends (Breiman, 2001). RF has the advantages of having more flexibility and fewer assumptions than traditional regression (e.g.

data distributions; Cutler et al., 2007); although, the black box method can be more difficult to interpret ecologically without providing direct estimates of trends.

We compare these two statistical methods in analyzing Kansas stream survey data for Topeka Shiner (*Notropis topeka*), a federally endangered and state-threatened minnow species. Multiple state agencies classify the Topeka Shiner as a threatened or endangered species (IDNR, 2015; KDWP, 2022). State wildlife action plans document substantial reductions in Topeka Shiner populations across their range (MNDR, 2015; Schneider et al., 2011; SDDGFP, 2023). Researchers and managers have accumulated data on Topeka Shiner since implementing protections in these states. Clearer and more consistent ecological insights from the existing field distribution of this uncommon fish could assist researchers and managers in identifying next steps for state-level conservation planning. Our purpose in integrating multiple analyses is to identify consistent and divergent ecological trends and guide future research and management.

Different studies of the same taxa can address the same primary conservation questions and result in different conclusions based on variation in field methods, location, or statistical analyses. As examples, studies across different state and watershed systems reveal variation in the needs of this fish for water quality (Galic et al., 2023; Simpson et al., 2019), system size (USFWS, 1997; Winston, 2002), and flow permanence (Hatch, 2001; Wall et al., 2004). In addition, general prairie stream fish studies differ in conclusions on whether native fish (including Topeka Shiner) are adversely affected by anthropogenic barriers (Barbarossa et al., 2020; Hitchman et al., 2018), land use changes (e.g., agriculture, grazing, urbanization; Gido et al., 2010; Yates & Bailey, 2006), and restructuring of the biotic community (Gozlan et al., 2010; Mitchell & Knouft, 2009). Furthermore, studies in Kansas show different correlates of Topeka Shiner presence than systems in other states with respect to dams (Schrank et al., 2001), stream

substrates (Mammoliti, 2014; Witte et al., 2009), and community assemblages (Campbell et al., 2016).

We address the challenge of using monitoring data with Topeka Shiner conservation as an example through three objectives. First, we analyze fish distributions and habitat needs from Kansas stream surveys using two statistical analysis methods (MLR, RF). Second, we integrate statistical results across approaches using comparative weight of evidence. We expect to reveal both points of model agreement and disagreement in the importance of variables, direction of effects, and predicted patterns. Finally, we contextualize and reconcile trends from our analyses with existing literature to draw ecological insights into the threatened and endangered Topeka Shiner and steer biologists to future research and management activities.

Methods

Study Organism

Topeka Shiner is a small-bodied (< 75mm), short-lived (< 3 years) minnow that epitomizes ecological and life-history characteristics of many prairie stream fishes (Hatch, 2001). This member of the family Cyprinidae occurs in Iowa, Kansas, Minnesota, Missouri, Nebraska, and South Dakota (Mammoliti, 2014). Topeka Shiner is a species in need of conservation which was listed as federally endangered in 1998 (USFWS, 1997) and threatened in the state of Kansas in 1999. Even with its threatened status, a non-trivial amount of distributional data is available for this species in Kansas. Thus, with considerable available data and conservation need, this taxon is an ideal candidate to compare analyses of drivers and disturbances that can reduce native distributions.

Topeka Shiner ecology and habitat use is highly variable throughout its range. For example, the fish can occur in small, headwater streams with cool, groundwater-fed pools and coarse substrates (Kerns & Bonneau, 2002; USFWS, 1997; Winston, 2002), but in some systems they have been reported to persist in larger rivers, with warmer waters, higher turbidity, and finer sediments (Galic et al., 2023; Simpson et al., 2019; Wall & Berry, 2004). These fish can use gravel sunfish nests for spawning (Pfaff et al., 2023; USFWS, 1997) yet are often found in systems with low community complexity (Schrack et al., 2001; Stark et al., 2002). Topeka Shiner typically reaches sexual maturity at age one and survives for one spawning season (Hatch, 2001). Therefore, Topeka Shiner populations could have limited opportunity to recover from potentially adverse effects from groundwater depletion (Berg et al., 2004), dam construction (Schrack et al., 2001), land use conversion (Cross & Moss, 1987), predator introduction (Knight & Gido, 2005), and invasive competition in Kansas (USFWS, 1997).

Fish Data

Personnel from the Kansas Department of Wildlife and Parks (KDWP) have collected fish data since the 1970s. Here, we analyzed 10,624 unique stream sites across 12,086 total samples from all 12 major Kansas river basins. In the survey dataset, 81% of the fish data were collected between 2000 and 2021. We transformed fish abundances from research permits, historical collections, and standardized field sampling into presences and absences to accommodate variable gear types and sampling protocols. In our presence-absence dataset, Topeka Shiner was present at 341 sites across the Kansas-Lower Republican ($n = 180$), Neosho ($n = 127$), and Smoky Hill-Saline River basins ($n = 34$).

Environmental Data

At each fish sample site, we acquired data for 11 abiotic and biotic environmental variables from existing GIS datasets: (1) *stream depth*, (2) *curvature*, (3) *gradient*, (4) *baseflow index*, (5) *clay*, (6) *species richness*, (7) *grassland*, (8) *cropland*, (9) *pastureland*, (10) *development*, and (11) *number of dams* (Table 1.1). We assessed relevant Topeka Shiner literature to select variables which have been identified as potential drivers of the species' distribution (Rode et al., 2025). We removed fish sample records with missing environmental data, which reduced the number of Topeka Shiner presence sites ($n = 335$). Due to the large number of absence sites relative to presence sites, we balanced data to include all available presence sites and an equal number of randomly selected absence sample sites (Steen et al., 2021).

Multiple Logistic Regression (*Objective 1a*)

We used MLR as our first statistical approach following Mather et al. (2026) for its common use in ecology and straightforward interpretation. Specifically, we estimated simple, main effects of each of our 11 ecological variables on the binomially distributed response of Topeka Shiner presence or absence with a logit-link and maximum likelihood functions (Hosmer et al., 2013). Our 11-variable ecological model excluded highly correlated variables (Pearson's $P > 0.70$). We fit the MLR model using both original and standardized data with a Laplace estimator from the *glmmTMB* package in R (McGillucuddy et al., 2025). We show the MLR on original data (Appendix A Table 1) but focus our interpretation on the MLR that used standardized data (mean = 0; standard deviation = 1; Appendix A Table 2). We tested model

prediction power using 10-fold cross validation. From this, we measured accuracy in predicted presences with (1) confusion matrices and (2) area under the curve (AUC) calculated from a receiving operator characteristic (ROC) plot (*caret*, *pROC* packages; Kuhn & Max, 2008; Robin et al., 2011).

Data were collected independently at different sites and times (i.e. without repeated measures). We removed outliers and known erroneous data points (e.g. incorrect geographic locations, values beyond measurement scale). Modeling on the logit-transformed scale maintained the linear order of parameters. We inspected the model for multicollinearity among variables by estimating individual variable inflation factors (VIF) with the performance package (Lüdtke et al., 2021). We considered variables to be multicollinear if VIF scores exceeded 10 (Kutner, 2005). Thus, we met, assessed, or addressed the assumptions for independence, measurement error, linearity, and multicollinearity (Hosmer et al., 2013).

MLR Variable Importance

For MLR, we evaluated the importance of individual variables in contributing to the probability of Topeka Shiner presence using multiple analytical outputs: (1) statistical significance ($p \leq 0.05$), 95% confidence intervals; (2) relative size of parameter estimates (β ; standardized coefficients only), and (3) odds-ratios (OR) (Mather et al., 2026). First, we prioritized variables that differed from the null effect ($p \leq 0.05$ and a 95% CI that does not cross zero). Second, we assessed the relative impact of each significant variable on the log-odds likelihood of Topeka Shiner presence with the absolute magnitude of β estimates, which represent the change in probability of Topeka Shiner presence for every standardized unit change in the environmental variable. Third, we calculated OR from the standardized beta-coefficients.

Variables with $OR > 1$ indicated an increase in the odds of Topeka Shiner presence, < 1 indicated a decrease, and an odds ratio of 1 indicated no effect.

MLR Response Trends

We assessed trends between Topeka Shiner presence and each environmental variable. The positive or negative direction of estimated beta-coefficients indicated whether ecological variables were associated with increasing or decreasing likelihood of Topeka Shiner presence. We visualized marginal effects of individual variables by holding all other variables constant and transforming logit-scale responses to the predicted probability of Topeka Shiner presence.

Random Forest (*Objective 1b*)

For the second statistical approach, we used a random forest (RF) algorithm following Mather et al. (2026) for its growing popularity in ecology and flexibility in assumptions. To estimate the impact of the same ecological variables on the probability of Topeka Shiner, we used the same original dataset with simple effects. We tuned the hyperparameters *mtry* (maximum number of variables possible for the algorithm to test at each node) and *ntrees* (number of trees in the forest) using the *randomForest* package in R (Liaw & Wiener, 2002). Based on the method outlined in Faraway (2016), we set *mtry* to minimize error using cross-validation. In total, we bootstrapped 3000 trees (*ntrees*) to calculate estimates and error rates. We assessed model performance with the confusion matrix and out-of-bag error metric native to RF, deriving model accuracy (%) by subtracting the error rate from 100. We tested model prediction by producing ROC curves and evaluating the AUC (*pROC* package; Robin et al., 2011). The

assumptions for independence and measurement error were met or assessed as in the MLR model (Breiman, 2001).

RF Variable Importance

We evaluated the relative importance of variables with two metrics: (1) mean decrease in accuracy and (2) mean decrease in the Gini index. Mean decrease in accuracy quantified relative importance by excluding each variable and estimating the change in accuracy of final prediction classifications. Mean decrease in the Gini index measured relative importance by excluding each variable and estimating the change in purity of nodes within trees. For both indices, the greater the mean decrease in the accuracy or Gini index from excluding a variable, the more important the variable. For the 11 covariates, we visually assessed variable importance plots. Without objective criteria for RF thresholds, we selected a natural breakpoint in the highest relative importance as indicated by accuracy and Gini metrics.

RF Response Trends

We visualized the partial dependencies of individual variables on predicted probability of Topeka Shiner presence, while holding all other values at their mean. We converted responses from the default settings on the logit-scale to probabilities of fish presence and applied Locally Estimated Scatterplot Smoothing (LOESS) to raw predictions to generalize trends (Breiman, 2001; Cleveland, 1979).

Comparison and Integration Across Statistical Approaches (*Objective 2*)

The MLR and RF differ in outcomes (direct parameter estimates for β and OR versus averaged trends) and statistical metrics (p and CI versus accuracy and purity), and we currently lack standard practices in ecology and fisheries to directly compare across these approaches. Thus, we compared the two statistical models with three types of outputs as outlines in Mather et al. (2026): (1) variable importance ranks, (2) probability plots, and (3) spatial model predictions.

Variable Importance

To enable comparisons of environmental variable importance across the two models, we used a ranking approach. For simplicity, we selected one measure of relative importance for each approach, because each importance metric produced similar outcomes for interpretation. For MLR, we ordered variables by statistical significance then by decreasing magnitude of standardized estimates (e.g. $|-0.75| > |0.25|$). For RF, we used the order of variables already assigned by accuracy. By converting each metric of importance to relative ranks on the same scale (highest rank = 1), we integrated the dissimilar approaches of MLR and RF.

Response Trends

We compared the direction and shape of response trends for variables across the two approaches to evaluate the consistency of data signals for influences on Topeka Shiner presence. Point estimates were not available for RF modeling as in MLR. Therefore, we show plots of predicted probability of presence as a function of covariate values for both approaches.

Map Predictions

We applied predictions from both model approaches to the full Kansas dataset with *caret* in R (Kuhn & Max, 2008) to visually compare the resulting geospatial patterns in ArcGIS Pro (Esri Inc., 2024). The MLR produces continuous predictions from 0 to 100% probability of Topeka Shiner presence, while the RF model predicts the probability that samples are classified as fish presence (1) for each tree and generates a binary outcome of less or greater than 50% probability of presence among all trees by default. We chose to identify consistency among spatial patterns by querying individual predictions for low (below 50% MLR and 0 RF) or high (above 50% MLR and 1 RF) probability of fish presence within major watersheds.

Results

Multiple Logistic Regression (*Objective 1a*)

The MLR analysis identified significant influences of *clay* (95% CI [1.55, 2.59]; $p \leq 0.005$), *stream depth* (95% CI [-3.17, -0.56]; $p \leq 0.01$), *number of dams* (95% CI [-1.11, -0.47]; $p \leq 0.005$), and *species richness* (95% CI [-0.96, -0.41]; $p \leq 0.005$) on the predicted probability of Topeka Shiner presence (Table 1.2; Figure 1.2A-D). The remaining variables had non-significant relationships. Predicted probability of Topeka Shiner presence exhibited the strongest response to *clay* (Table 1.2; Figure 1.2A), as reflected by the largest standardized slope in our analysis ($\beta = 2.07$) and a nearly eight-fold increase in the chances of Topeka Shiner presence with one standardized unit increase in *clay* (OR = 7.91; SD = 12%). For *stream depth*, predicted probability of Topeka Shiner presence displayed the second strongest response (Table 1.2; Figure 1.2B), as indicated by the standardized slope ($\beta = -1.86$) and 84% reduction in the odds of Topeka Shiner presence as *stream depth* increases by one standardized unit (OR = 0.16; SD = 2.5 m; Figure 1.2B). Predicted probability of Topeka Shiner presence had the third strongest

response to *number of dams* based on the standardized slope ($\beta = -0.79$; Table 1.2; Figure 1.2C), and as the *number of dams* increased by one standardized unit (SD = 8 dams), the likelihood of Topeka Shiner presence reduced by half as (OR = 0.45). Finally, predicted probability of Topeka Shiner presence revealed the fourth strongest response to change in fish *species richness* (Table 1.2; Figure 1.2D), based on the standardized slope ($\beta = -0.69$) and a 50% reduction in odds that Topeka Shiner will be present as *species richness* increased by one standardized unit (OR = 0.50; SD = 8 species).

Random Forest (*Objective 1b*)

The RF analysis identified *species richness*, *clay*, and *baseflow index* as the highest ranking variables in terms of importance (i.e., accuracy and the Gini index; Figure 1.3A-B). *Species richness* was associated with the greatest mean decrease in accuracy (Figure 1.3A) and the second greatest mean decrease in the Gini index when excluded among trees (Figure 1.3B). The predicted probability of Topeka Shiner presence decreased with increasing *species richness* (Figure 1.4A). Trees that excluded *clay* among the RF analysis experienced the second greatest mean decrease in accuracy (Figure 1.3A) and the greatest mean decrease in the Gini index (Figure 1.3B). Increases in *clay* were associated with increased probability of Topeka Shiner presence (Figure 1.4B). Trees that excluded *baseflow index* had the third greatest mean decrease in accuracy (Figure 1.3A) and the third greatest mean decrease in the Gini index (Figure 1.3B). As indicated by partial dependence plots, the predicted probability of Topeka Shiner presence exhibited a nonlinear relationship with *baseflow index*. The predicted probabilities increased between *baseflow index* of 10 to 20, peaked between 20 – 30, and decreased for *baseflow index* higher than 30 (Figure 1.4C). Other variables (*curvature*, *development*, *gradient*, *number of*

dams, grassland, stream depth, cropland, and pastureland) contributed to relatively minor changes in RF accuracy and the Gini index (Figure 1.3).

Comparisons (*Objective 2*)

Our MLR and RF models performed well on all evaluation metrics (Figure 1.5). The MLR model was 88% accurate and 93% precise; there were fewer false presences than false absences (Figure 1.5A). With 7.16% out-of-bag error, our RF model was 93% accurate with a precision of 93%; the model predicted more false presences than absences (Figure 1.5B). Additionally, both models produced high sensitivity and specificity regarding the AUC (MLR: 0.935; RF: 0.998; Figure 1.5C-D).

Variable Importance

The MLR and RF analyses agreed on the importance of *clay* and *species richness* (Table 1.3). *Clay* ranked #1 for MLR variable importance, ranked #1 and #2 in importance for RF accuracy and the Gini index (Table 1.3). *Species richness* ranked #4 in MLR variable importance, ranked #1 and #2 for the two RF measures (Table 1.3).

The relative effects of *stream depth, number of dams, baseflow index* on the probability of Topeka Shiner presence differed between the MLR and RF analyses (Table 1.3). *Stream depth* had a very high importance ranking and was statistically significant in the MLR (rank # 2) but had a very low importance ranking for the RF analysis (ranks #9, #9; Table 1.3). Similarly, *number of dams* was statistically significant and influential in the MLR (rank #3) but had lower influence ranks in the RF analysis (ranks #6, #7; Table 1.3). *Baseflow index* was not significant

in the MLR and ranked low in importance (rank #8), but it received high importance rankings for RF (ranks #3, #3; Table 1.3).

Response Trends

The MLR and RF analyses agreed on the response trends of *clay*, *species richness*, *stream depth*, and *number of dams* (Figure 1.6; Figure 1.7). *Clay* showed rapidly increasing probability of Topeka Shiner presence between *clay* values of 20% and 40% (Figure 1.6A-B). Across models, *species richness* showed the same trend of decreasing predicted probability of Topeka Shiner presence as *species richness* increased (Figure 1.6C-D). Both statistical approaches showed similar trends for predicted probability of Topeka Shiner presence decreasing as *stream depth* increased (Figure 1.7A, B). Similarly, plots showed decreases in predicted probability of Topeka Shiner as the *number of dams* increased (Figure 1.7C-D).

The MLR and RF analyses disagreed on the response trend of *baseflow index*. The MLR predictions for *baseflow index* exhibited a shallow, negative, highly variable relationship with probability of Topeka Shiner presence (Figure 1.7E). Meanwhile, visualizations of RF model predictions showed a less variable, nonlinear response with a peak > 50% probability of Topeka Shiner presence between *baseflow index* values of 20-30 (Figure 1.7F).

Spatial Trends

The predicted spatial trends from both statistical methods also presented points of agreement and disagreement (Figure 1.8). For both MLR and RF models, areas of high (> 50%) predicted probability of Topeka Shiner presence clustered in the Flint Hills region (EPA Ecoregion Level 3) of central Kansas (Figure 1.8), overlapping with portions of the Kansas-

Lower Republican, Neosho, and Smoky Hill-Saline River basins (Figure 1.1A). While high probability of presence concentrated in the Kansas-Lower Republican watersheds, sites with predicted high probabilities also occurred across the Neosho River basin and as far as the western border in the Smoky Hill-Saline River basin (Figure 1.1A; Figure 1.8). Some sites presented disagreements between the two models. For example, the MLR model predicted more sites in the Smoky Hill-Saline River Basin to have higher probabilities (Figure 1.1A; Figure 1.8A), whereas the RF model predicted more sites spanning the Neosho River Basin to have higher probabilities of Topeka Shiner presence (Figure 1.1A; Figure 1.8B).

Discussion

Agreement in Methodological Approaches Can Build Confidence (*Objectives 1, 2, 3*)

Identifying agreement in quantitative trends across statistical analyses can strengthen data-based ecological inferences extracted from long-term monitoring datasets. Given the unclear, ambiguous, or contradictory results that can emerge from stand-alone statistical analyses of field data, identifying consistent outputs with different methods offers one way to address underlying uncertainty (Johnson, 2002; Mather et al., 2026). Some investigators in marine (Becker et al., 2020; Hazen et al., 2021; Hebert et al., 2024) and terrestrial systems (O'Malley et al., 2024; Shoemaker et al., 2018) similarly identify robust ecological relationships for other taxa by comparing consistent trends across multiple statistical analyses. Although, most studies of species distributions do not systematically compare consistency of trends across duplicate analyses. Our cross-analysis comparisons of variables and spatial predictions confirmed consistent and influential roles of *clay* and *species richness* on Topeka Shiner presence. Additionally, we identified locations with consistently high predicted probability of presence that

offer high-value locations for focused future sampling and hypothesis testing where Topeka Shiner populations are more likely present.

Compared to our paired statistical methods, most ecological studies of Topeka Shiner report habitat relationships and predictions of species distributions from a single type of statistical analyses (e.g. Schrank et al., 2001; Wall et al., 2004). For example, several single-method studies attribute the importance of stream substrates (e.g. gravel) to how they affect fish feeding, spawning, and predator evasion behaviors of Topeka Shiner (Bakevich et al., 2015; Witte et al., 2009). By comparing outputs of our two different statistical analyses, we identified a consistent indicator that *clay* soil composition in the catchment area could influence Topeka Shiner presence. Presenting a strong positive relationship between Topeka Shiner and *clay* from multiple analyses establishes the need for further examination of this surprising trend. We learned from this ecological insight that existing measurements of *clay* composition at the stream site level may help researchers and managers better understand why Topeka Shiner are present or absent at specific sites.

Our comparison across two statistical analyses confirmed a consistent and influential relationship between Topeka Shiner and the biotic fish community, specifically that more diverse fish communities with higher *species richness* adversely affected Topeka Shiner. The local effect of biota on Topeka Shiner varies across other studies. For example, total *species richness* exhibits a positive relationship with Topeka Shiner in oxbows (Simpson et al., 2019) and negative effects in stream pools (Schrank et al., 2001). At the statewide level among streams, our models present stable correlations between Topeka Shiner presence and community diversity with consistent thresholds regardless of statistical method. Thus, our process for identifying agreement across statistical analyses points to prevailing global effects rather than context-

dependent relationships of Topeka Shiner with *species richness*. From this, we learned that *species richness* could help conservation professionals identify which sites are likely to support Topeka Shiner reintroductions.

Disagreement in Statistical Outcomes for Ecological Drivers of Distribution Guides Future Work (*Objectives 1, 2, 3*)

While disagreement across model results generates uncertainty and can lead to confusion, disagreement across statistical approaches can also suggest new directions for collecting and analyzing field data for native organisms facing threats with environmental change. Some freshwater fish ecology studies compare models to improve accuracy (e.g. Olden & Jackson, 2002) or advance interpretability (e.g. Chollet Ramampandra et al., 2023). However, these comparisons are more often made to prescribe one approach than to retain ecological information or identify disagreement among trends across approaches, as we do here (but see Shoemaker et al., 2018, terrestrial; Huang & Frimpong, 2015, aquatic). Analyzing disagreements among statistical analyses is a novel way to extract ecological insights from field survey data that can guide next steps for future conservation actions (Mather et al., 2026). In our research, identifying inconsistencies in the relative magnitude, direction, and shape of relationships for *stream depth*, *number of dams*, and *baseflow index* points to key knowledge gaps. Furthermore, examining diverging model predictions for specific sites in the Missouri, Lower Arkansas, and Walnut River basins can identify priority locations for examining these conflicting trends with consistent sampling across the range of inconsistent variables.

For *stream depth*, disagreements between MLR and RF hint that additional system size and habitat metrics may be necessary to understand the importance of *stream depth* for Topeka

Shiner. Topeka Shiner is often associated with small streams (Stark et al., 2002; USFWS, 1997; Winston, 2002). However, Topeka Shiner may (Aadland & Kuitunen, n.d.; Galic et al., 2023) or may not (Schrank et al., 2001) use varying *stream depths*. The different relative effects of *stream depth* in our two models of Topeka Shiner presence could be attributed to RF detecting nonlinear, interactive, or local effects while MLR assumes a consistently linear, global effect. This insight aligns with the idea that *stream depth* could influence the species on multiple habitat scales (Johnson, 1980). With this in mind, conservation professionals could select future sample sites with varying stream depths and in-situ habitat depths to further test which conditions may best support restoration of Topeka Shiner populations.

For the *number of dams*, effect size disagreements identified a need to fill gaps in data coverage at the state-wide scale. Ecologists often discuss the influences of dam-related fragmentation (Barbarossa et al., 2020), sedimentation (Dodds et al., 2004), and flow regime alteration on stream fish communities (Poff et al., 1997; Ward et al., 1983). However, Topeka Shiner studies focus on negative relationships with small farm impoundments (Gerken & Paukert, 2013; Schrank et al., 2001), whereas our dataset only incorporated dams of greater heights and storage capacities (7.62 m with 1.85 ha-m, or 1.83 m with 6.17 ha-m; Cooper & Infante, 2024). Our results indicated consistently negative responses of Topeka Shiner to increasing the total *number of dams* (irrespective of function or position in the stream). Although, the disagreement in relative effect across analyses may be attributed to the MLR presenting to stable coefficients where the RF detected an interrupted pattern. Without validation using complementary methods, an incomplete picture for dam data could have been interpreted erroneously as extremely or minimally important to Topeka Shiner habitat. Researcher-manager teams could collect future monitoring data to target streams with intermediate (5-15) *numbers of*

dams to fill the observed gap and solidify the strength of this relationship with large impoundments.

For *baseflow index*, the disagreement in the effect size, direction, and shape of trends illustrates competing ecological relationships (linear or non-linear) for linking Topeka Shiner with sustained water availability. Several studies examine in-stream flow effects on Topeka Shiner regarding flow endurance (Aadland & Kuitunen, n.d.; Adams et al., 2000; Bouska & Paukert, 2010) and permanence (Wall et al., 2004). However, relationships with subsurface flow are less often analyzed (Bayless et al., 2003; Wall et al., 2004). Early observations associate the Topeka Shiner with groundwater springs for permanent flow refugia (Mammoliti, 2014; USFWS, 1997; Wall & Berry, 2004). Although, Kansas has limited data for groundwater spring locations and contribution to surface-flow, which constrains direct investigation of these relationships. While not directly equivalent, our coarse metric of *baseflow index* relates to groundwater flow and other sustained sources of stream flow between runoff events, which can be highly influential and sometimes can have a nonlinear effect on other fishes (Huang & Frimpong, 2015). The conflicting trends for the influence of *baseflow index* on Topeka Shiner presence across models could be attributed to needing *a priori* specification of nonlinear trends in MLR whereas RF captures the more complex relationships by design. By using two modeling approaches, we utilized the strengths of each to identify potentially nonlinear trends that could have been overlooked using one model alone. Our findings suggest that analysts should consider examining a non-linear relationship between *baseflow index* and Topeka Shiner elsewhere to identify promising sites for future restoration efforts.

Integrating Approaches (*Objectives 2, 3*)

Comparing statistical approaches provides benefits, as described above, but also presents challenges for implementation. The MLR approach provides valuable metrics that quantify parameter estimates (β), amount of variation (95% CI), and effect size (β ; OR). Meanwhile, RF offers relative importance measures shaped by repeated classification can be a powerful approach to prediction (Cutler et al., 2007). While models are fitted to the same presence-absence data, the outputs are not directly comparable. Following the approach of Mather et al. (2026), we translated outputs from both analyses to a common scale by assigning relative ranks to each variable. These ranks allowed us to combine the interpretability of MLR with the predictive flexibility of RF.

Our dual-approach comparison provides the foundation for an integrated ecological summary of what is known and unknown about Topeka Shiner distributions. First, our analyses confirmed that Topeka Shiner typically occupies smaller streams (here measured as *stream depth* at the downstream reach). Other investigators have also found Topeka Shiner generally inhabits smaller streams in Kansas (Mammoliti, 2014; Winston, 2002), with shallower depths (Minckley & Cross, 1959; Stark et al., 2002). However, when investigators measure *stream depth* in oxbows (Galic et al., 2023), area in ponds and watersheds (Pfaff et al., 2023; Wahl et al., 2024), and position in stream networks (Wall et al., 2004), conflicting conclusions can emerge about what system size is most often associated with Topeka Shiner presence. Our integration across analyses and studies demonstrates need for more comprehensive morphometric measurements in streams where Topeka Shiner occurs. Integrating *stream depth* and other stream size metrics

across sites could lead to a more focused and consistent understanding of Topeka Shiner habitat within and across states.

Second, the integration of our two analyses shows that Topeka Shiner typically persists in fish communities with lower *species richness*. Our results reinforce some previous community diversity trends with Topeka Shiner populations in Kansas (Mammoliti, 2014; Schrank et al., 2001). Although, community composition may additionally influence Topeka Shiner, such as the relative abundance of individual associates and competitors (e.g. Simpson et al., 2019, 2021) or presence or density of stocked predators (Knight & Gido, 2005; Pfaff et al., 2023; Schmolke et al., 2019). As a future research direction from our model comparisons, systematic analyses of community composition (individual species or groups) may clarify the influences of community size on Topeka Shiner presence and identify data collection needs for biota in Kansas and across states.

Third, our analyses suggest that Topeka Shiner is more likely to occur in streams with fewer dams. Thus, our integrated results reinforced often speculated negative influences of dams on Topeka Shiner in Kansas (Mammoliti, 2014). However, confusion emerges when dam impacts are variably quantified as *number of dams* (our results), proportional area in watersheds, and other metrics within farm pond impoundments (Pfaff et al., 2023) and along streams (Gerken & Paukert, 2013; Schrank et al., 2001). More comprehensive assessment of the dam-relevant metrics (number, density, distance, size, etc.) and potential mechanisms influencing Topeka Shiner habitat (fragmentation, altered flow, predator introductions, etc.) could help identify and connect the roles of human barrier impacts on the fish (instream and off channel) that can help conservation professions prioritize needs related to dams.

Finally, in Kansas, *clay* and *baseflow index* emerged as potentially novel influences on Topeka Shiner. Upon further testing of relationship shapes for *baseflow index*, these broad-scale conditions may exert previously unrecognized influence on Topeka Shiner habitat. Generally, studies do not evaluate the role of non-streambed geology or subsurface hydrology on fish, but Wall et al. (2004) has suggested that bedrock types and groundwater potential at the catchment scale could affect Topeka Shiner in South Dakota. Based on the moisture-retention properties of clay soil (Dingman, 2015), increased *clay* composition in surface geology may buffer streams from water loss and support baseflows (Price, 2011). We propose that catchment-level *clay* composition and *baseflow index* may enhance the survival of Topeka Shiner and other prairie fish during stressful low water events. Our results present a novel idea that both soil and groundwater conditions may alleviate hydrological stress on Topeka Shiner populations in Kansas. Our approach revealed a new research direction for the Topeka Shiner data collection and analysis, especially in the face of intensifying drought and flood cycles related to climate change (Dodds et al., 2004). Specifically, additional data collection and analyses could parse the influence of *clay* and *baseflow index* on Topeka Shiner presence further. Field validation may clarify for conservation professionals which locations will not support Topeka Shiner, could support Topeka Shiner with habitat modification or restoration, and currently support natural populations that need close monitoring of habitat function.

Conclusions

The state agencies responsible for reviewing and mitigating ecological impacts on threatened and endangered species have finite time and resources to provide data-based, site-specific recommendations relevant to land-use, water allocation, and other proposed

development. Our study addresses the need to inform future resource allocation for conservation research and management by using a data-driven plan to circumvent persisting statistical challenges of drawing ecological insights from available monitoring data. Statistically comparing similar and diverging results can offer a diagnostic lens for models of natural field patterns and aggregate data-based evidence to coordinate future data collection and analyses. Together, agreements and disagreements between our MLR and RF analyses identified priority areas for further investigation for Topeka Shiner. In our study, agreement strengthened our confidence that we had appropriately modeled *clay* and *species richness*. Furthermore, disagreements provided informative opportunities to guide further research and field testing for *stream depth*, *number of dams*, and *baseflow index*. Compatible with the RAD management framework, our statewide assessment identifies some sites with consistently low potential for Topeka Shiner presence (both historical and present), for which conservation professionals may accept the current absent status of Topeka Shiner (Rahel & Lynch, 2022). Additionally, our comparative analyses can help identify priority variables for habitat management intervention (i.e. resist or direct). By interpreting complementary analyses together, our process provides a useful first step in distinguishing clear ecological trends from more ambiguous patterns and integrating pieces of data-based evidence to support confident, management-relevant decisions.

Acknowledgments

The project was administered by the Kansas Cooperative Fish and Wildlife Research Unit as a joint effort between Kansas State University, the U.S. Geological Survey, U.S. Fish and Wildlife Service, the Kansas Department of Wildlife and Parks (KDWP), and the Wildlife

Management Institute. We thank KDWP for sharing the monitoring dataset. We thank Michael Madin and Jean Francois from the Kansas State Department of Geography and Geospatial Analysis who have offered their expertise in R programming and spatial data analysis. Jordan Hoffmeier provided valuable insights into conservation concerns highlighted within the Kansas SWAP as well as substantial expertise in aquatic ecology. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. A review of scientific and common names was completed. Tribe and geologic names do not apply to this manuscript.

Tables

Table 1.1. Eleven ecological variables used to model Topeka Shiner distributions in Kansas. Included are variable name, relevance to fish, definition of measurement, units, spatial scale, database used, and source.

Variable Name	Fish Relevance	Definition	Units	Spatial Scale	Database	Source
<i>Stream Depth</i>	The depth of streams influences flow regimes and structures fish communities. Shallow streams are associated with less flow and smaller communities, while deeper areas have larger quantities of flow and fish species.	Estimated depth for flowline exit point.	Meters	NHDPlus Flowlines	National Hydrography Data Plus V2 (NHDPlus)	(McKay et al., 2012)
<i>Curvature</i>	Curvature is a measure of stream channel sinuosity and is associated with microhabitat diversity. Channelization, flow alteration, and bank stabilization for human use reduces sinuosity and straightens river paths.	Calculated average geometric curvature from flowline points at least 60m apart.	1/Meters	NHDPlus Flowlines	NHDPlus	(McKay et al., 2012)
<i>Gradient</i>	Channel slope, or gradient, is a commonly measured fish habitat variable. A reaches gradient can dictate how quickly the river flows, and therefore what habitat types and fauna are present.	Slope of flowline, calculated as the change in elevation over the length of the flowline.	Meters/Meters	NHDPlus Flowlines	NHDPlus	(McKay et al., 2012)
<i>Baseflow Index</i>	Baseflow is the quantity of water in stream channels excluding precipitation runoff. Human-altered flow regimes can reduce baseflow, affecting critical ecosystem function for stream fish.	Calculated contribution of groundwater and other sustained sources to total stream flow	%	NHDPlus Catchments	The Stream Catchment Database (StreamCat)	(Hill et al., 2016)

Table 1.1. Eleven ecological variables used to model Topeka Shiner distributions in Kansas. Included are variable name, relevance to fish, definition of measurement, units, spatial scale, database used, and source.

Variable Name	Fish Relevance	Definition	Units	Spatial Scale	Database	Source
<i>Clay</i>	Soil composition of the stream channel and adjacent catchment areas may influence bed substrate and regulate runoff and baseflow delivery to streams.	Calculated average Soil Surface Horizon A Composition, ratio of mass of clay (<0.002mm) to mass of soil fines.	%	NHDPlus Catchments	Soil Survey Geographic Database (SSURGO)	(NRCS, 2022)
<i>Species Richness</i>	Species richness is commonly used as a measure of ecosystem health and biotic community size.	Calculated from all fish species present in one stream sample	#	Site Sample	Kansas Department of Wildlife & Parks (KDWP)	n.p.
<i>Grassland</i>	Grassland cover is associated with prairie and plains ecosystems. Kansas' predominately historic grasslands are associated with co-evolved stream fish communities.	Calculated grassland cover in 2021	%	NHDPlus Catchments	National Land Cover Database (NLCD)	(Dewitz, 2023)
<i>Cropland</i>	Agricultural practices, including land used for crop production, can potentially alter nearby stream systems through water extraction or sediment runoff. Stream habitat alterations (flow, turbidity, etc.) can lead to declines in fish populations.	Calculated agricultural cropping land use cover in 2021	%	NHDPlus Catchments	NLCD	(Dewitz, 2023)
<i>Pastureland</i>	Agricultural practices, including land used grazing or haying, can potentially alter nearby stream systems through nutrient runoff. Stream habitat alterations (productivity, algal blooms, etc.) can lead to declines in fish populations.	Calculated agricultural pastureland use cover in 2021	%	NHDPlus Catchments	NLCD	(Dewitz, 2023)

Table 1.1. Eleven ecological variables used to model Topeka Shiner distributions in Kansas. Included are variable name, relevance to fish, definition of measurement, units, spatial scale, database used, and source.

Variable Name	Fish Relevance	Definition	Units	Spatial Scale	Database	Source
<i>Development</i>	Human development such as urbanization can lead to changes in water quality and channel structure of streams. Therefore, developed land use of varying intensities could alter fish distribution.	Calculated developed land use cover (Open space, low-high intensities) for 2021	%	NHDPlus Catchments	NLCD	(Dewitz, 2023)
<i>Number of Dams</i>	Human construction of dams can alter the stream habitat, water quality, and system connectivity. Dams can change fish assemblages by providing more lentic habitat for some fish and removing lotic habitat for others.	Total mainstem dam count (>7.62 m with 1.85 ha-m of storage, or 1.83 m with 6.17 ha-m of storage)	#	NHDPlus Flowlines	National Anthropogenic Barriers Dataset (NABD) & NHDPlus	(Cooper & Infante, 2024; Ostroff et al., 2013)

Table 1.2. Results from multiple logistic regression on standardized data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas from 1947 to 2020 ($n=670$). Included are variable name, coefficient (β), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p -value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor variables include *clay* (% in catchment surface geology), *stream depth* (estimated in meters), *number of dams* (total in the mainstem), *species richness* (number of fish species in a single sample), *development* (% cover in the catchment), *grassland* (% cover in catchment), *pastureland* (% cover in the catchment), *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *curvature*, *cropland* (% cover in the catchment), and *gradient* (m/m).

Variable	β	SE	Z	95% CI		p		OR	VIF
				Lower	Upper				
(Intercept)	-1.35	0.38	-3.52	-2.10	-0.60	0.00	*	0.26	
<i>Clay</i>	2.07	0.26	7.84	1.55	2.59	0.00	*	7.91	1.30
<i>Stream Depth</i>	-1.86	0.67	-2.80	-3.17	-0.56	0.01	*	0.16	1.77
<i>Number of Dams</i>	-0.79	0.16	-4.84	-1.11	-0.47	0.00	*	0.45	1.78
<i>Species Richness</i>	-0.69	0.14	-4.87	-0.96	-0.41	0.00	*	0.50	1.28
<i>Development</i>	-1.08	0.66	-1.64	-2.37	0.21	0.10		0.34	1.38
<i>Grassland</i>	0.72	0.41	1.75	-0.09	1.54	0.08		2.06	9.91
<i>Pastureland</i>	-0.24	0.22	-1.13	-0.67	0.18	0.26		0.78	2.84
<i>Baseflow Index</i>	-0.24	0.20	-1.18	-0.63	0.16	0.24		0.79	1.30
<i>Curvature</i>	0.12	0.19	0.62	-0.25	0.49	0.54		1.12	1.17
<i>Cropland</i>	0.09	0.29	0.31	-0.48	0.66	0.75		1.09	6.31
<i>Gradient</i>	0.03	0.16	0.17	-0.29	0.35	0.87		1.03	2.16

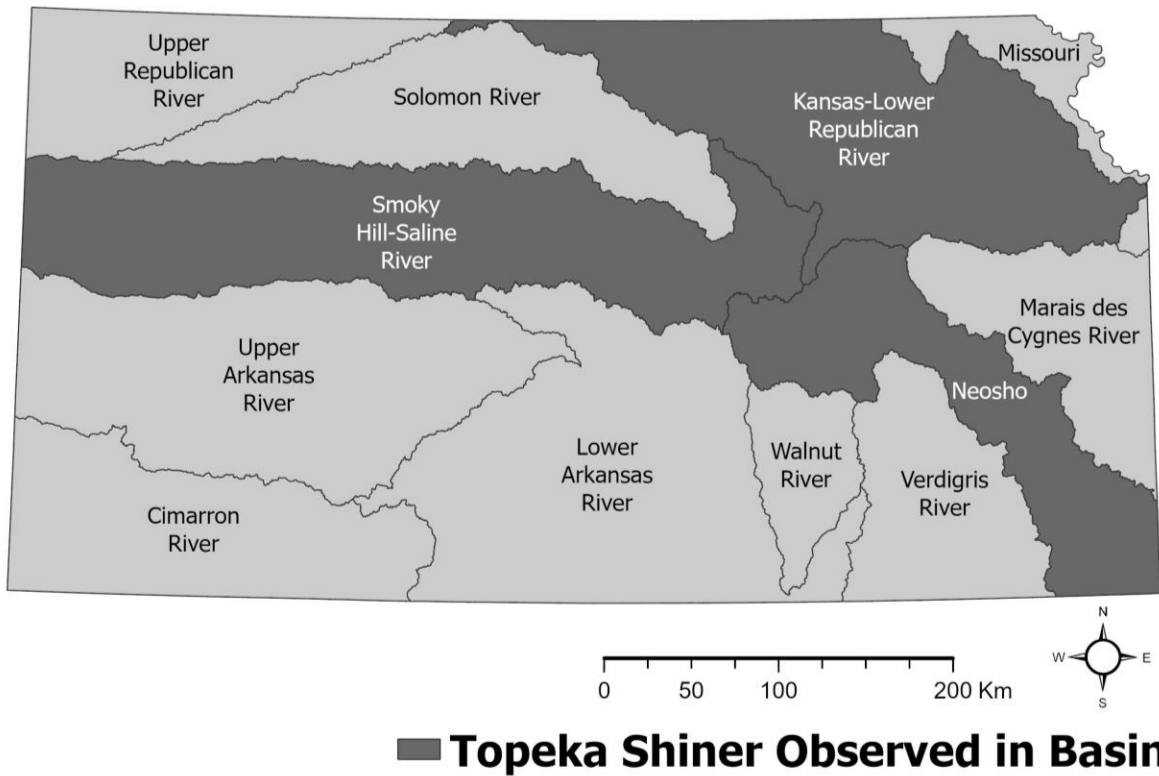
Table 1.3. Comparison of importance (ranks 1-11 from most to least important) for 11 ecological variables as identified by significance and the absolute magnitude of multiple logistic regression (MLR) estimates and mean decrease in accuracy of random forest (RF) models. Significant MLR estimates ($p \leq .05$) are denoted with *.

Variable	MLR (β)	RF (Accuracy)
<i>Clay</i>	1*	2
<i>Stream Depth</i>	2*	9
<i>Number of Dams</i>	3*	7
<i>Species Richness</i>	4*	1
<i>Development</i>	5	5
<i>Grassland</i>	6	8
<i>Pastureland</i>	7	11
<i>Baseflow index</i>	8	3
<i>Curvature</i>	9	4
<i>Cropland</i>	10	10
<i>Gradient</i>	11	6

Figures

Figure 1.1. Distribution of Topeka Shiner presences in stream assessment samples among (A) major river basins in (B) Kansas, USA.

A. Major River Basins in KS



B. Kansas, USA



Figure 1.2. Marginal effects on the predicted probability of Topeka Shiner presence (Y axis) from the top multiple logistic regression model that investigates 11 ecological variables. Predictor ecological variables on the X axis include: (A) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (B) *stream depth* (estimated in meters), (C) *clay* (% in catchment surface geology), (D) *number of dams* (total in the mainstem), (E) *species richness* (number of fish species in a single sample), (F) *gradient* (m/m), (G) *grassland* (% cover in catchment), (H) *cropland* (% cover in the catchment), (I) *curvature*, (J) *development* (% cover in the catchment), and (K) *pastureland* (% cover in the catchment). Raw data distributions are depicted with black points at the top and bottom of each panel, response trends are black lines, and confidence intervals are shown in grey. The model results in this figure correspond to Table 2.

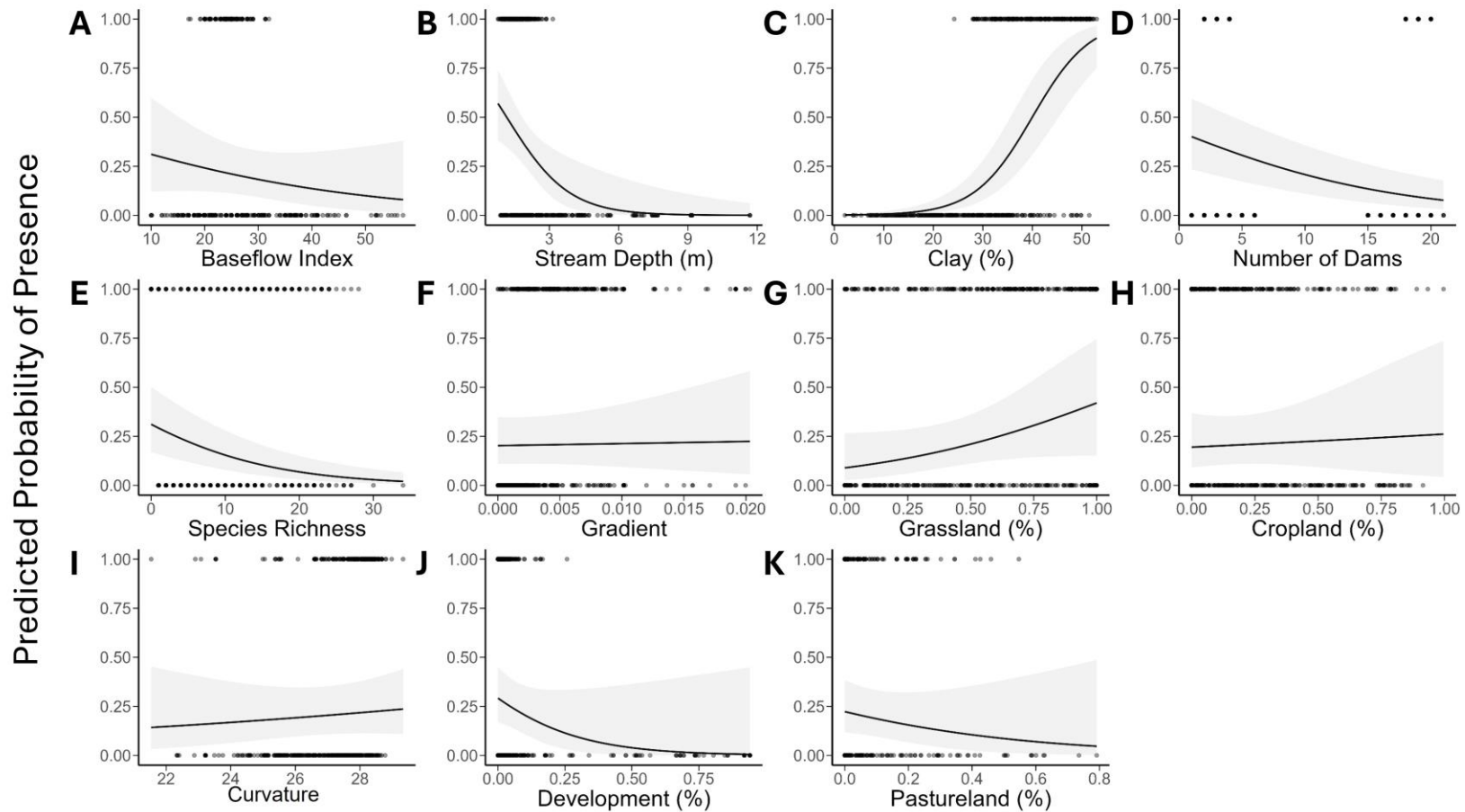


Figure 1.3. Importance of 11 ecological variables describing Topeka Shiner population distributions in Kansas from 670 stream samples collected between 1947 and 2020 using (A) mean decrease in accuracy and (B) mean decrease in Gini purity index modeled by random forest on the X axes. Predictor ecological variables on the Y axes include: *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *stream depth* (estimated in meters), *clay* (% in catchment surface geology), *number of dams* (total in the mainstem), *species richness* (number of fish species in a single sample), *gradient* (m/m), *grassland* (% cover in catchment), *cropland* (% cover in the catchment), *curvature*, *development* (% cover in the catchment), and *pastureland* (% cover in the catchment).

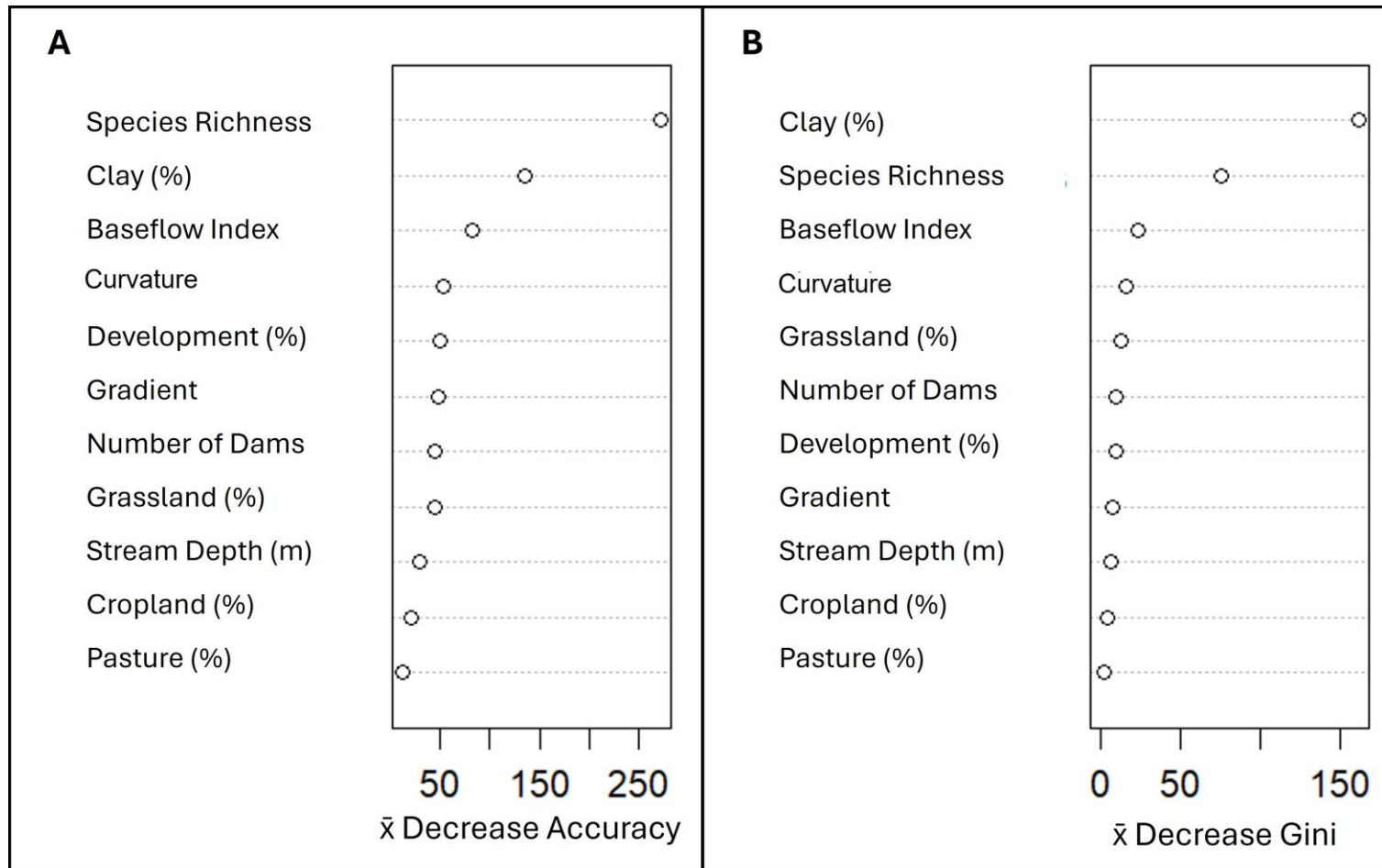


Figure 1.4. Partial Dependency of predicted probability of Topeka Shiner presence from 11 ecological variables. Predictor ecological variables include: (A) *species richness* (number of fish species in a single sample), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *curvature*, (E) *development* (% cover in the catchment), (F) *gradient* (m/m), (G) *number of dams* (total in the mainstem), (H) *grassland* (% cover in catchment), (I) *stream depth* (estimated in meters), (J) *cropland* (% cover in the catchment), and (K) *pastureland* (% cover in the catchment). Raw data distributions are shown as black points, raw response trends as black lines, and LOESS smoothed trends with confidence intervals in grey. (B)

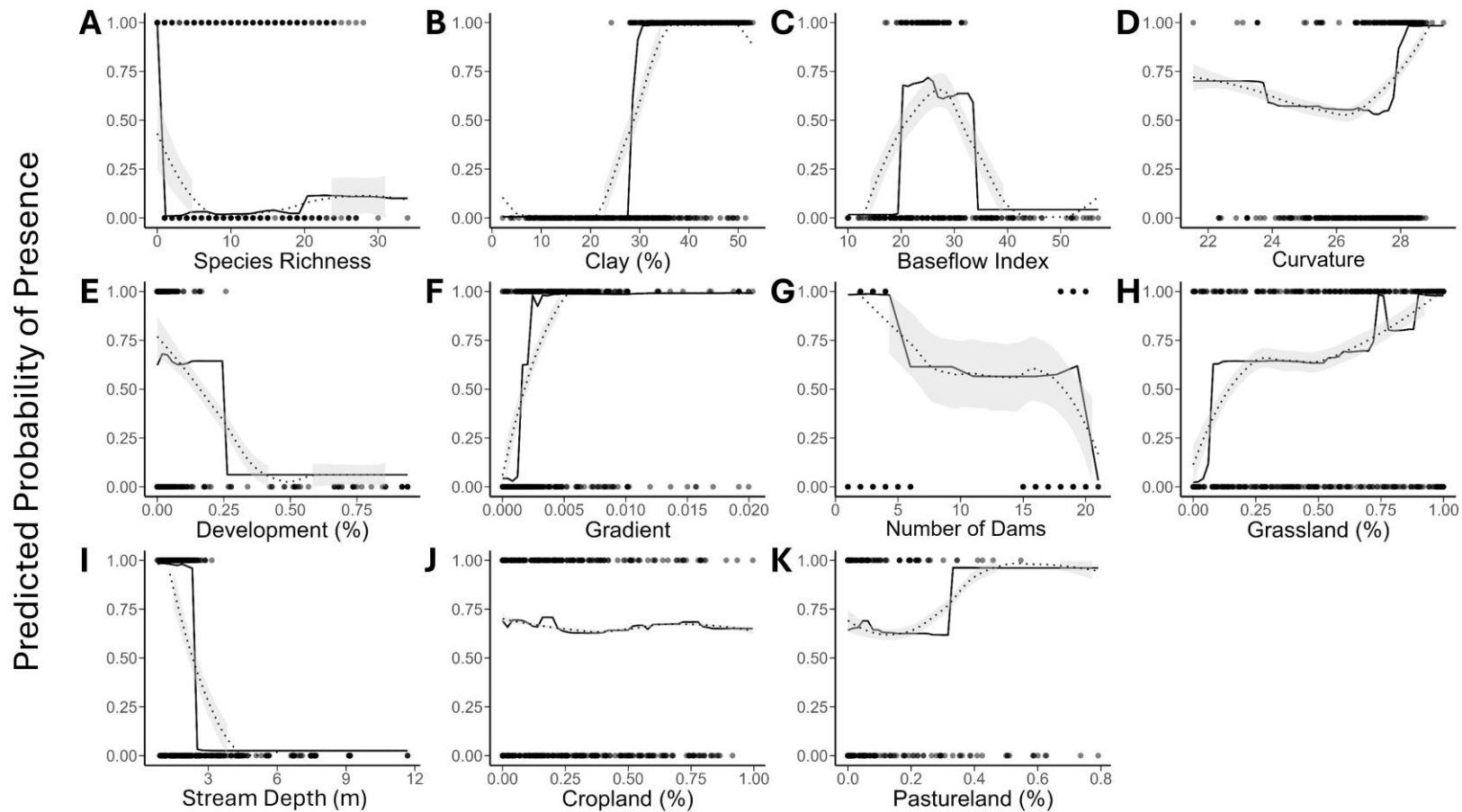


Figure 1.5. Multiple logistic regression and random forest model performance as described by (A), (C) confusion matrices with predicted values on the X axis and observed values on the Y axis; and (B), (D) receiving operator characteristic (ROC) curves with 1-specificity on the X axis, sensitivity on the Y axis, model predictions in solid, and expected values in dashed lines. Multiple logistic regression accuracy = 88.21%. Area under the curve (AUC) = 0.935. Random forest out-of-bag error = 7.16%. Accuracy = 92.84%. AUC = 0.998.

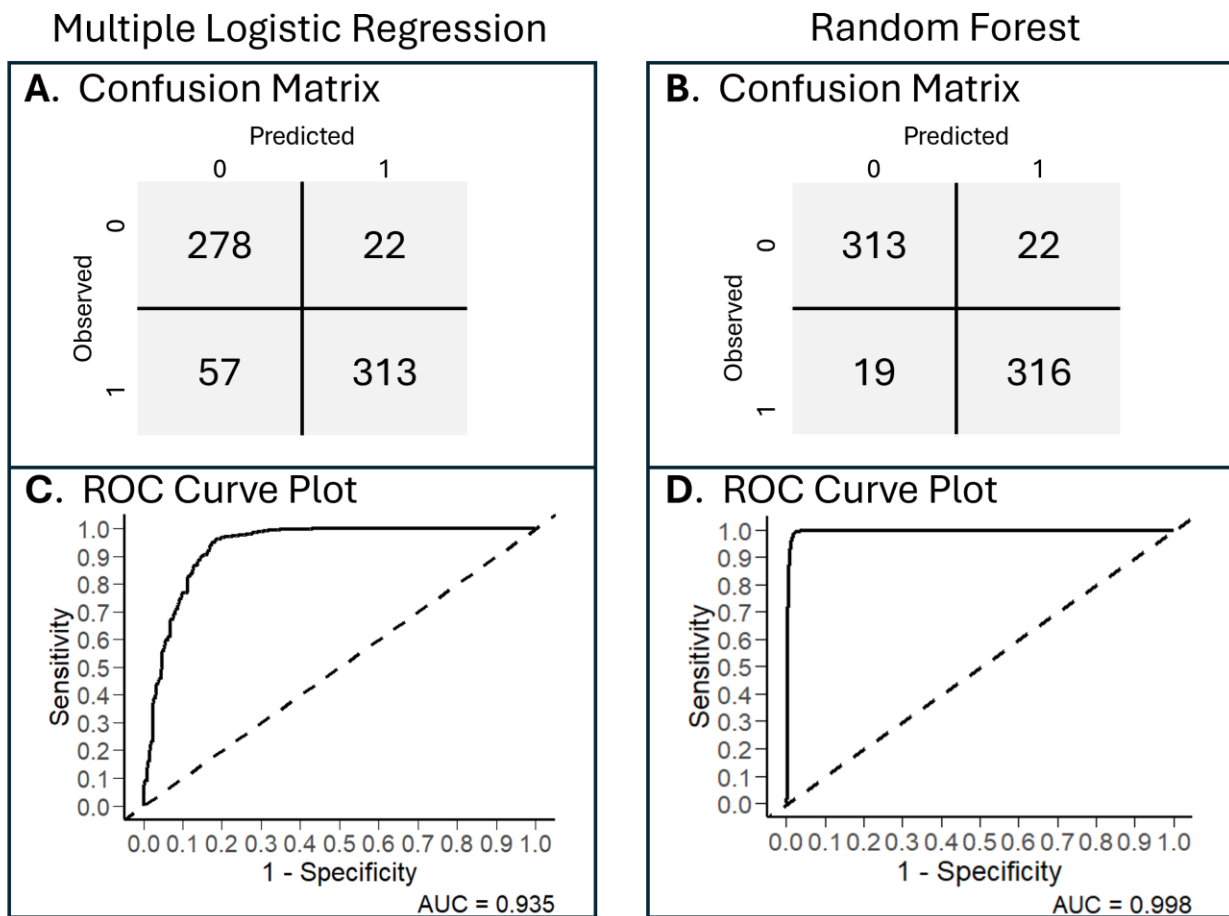


Figure 1.6. Predicted probability of presence in response to clay substrate composition (%) and species richness across the state of KS from (A), (C) multiple logistic regression and (B), (D) random forest ecological models of Topeka Shiner distributions in Kansas from 670 stream samples collected between 1947 and 2020. Predictor ecological variables include: (A), (B) *clay* (% in catchment surface geology) and (C), (D) *species richness* (number of fish species in a single sample) on the X axes. Response on the Y axes represents predicted probability of Topeka Shiner presence.

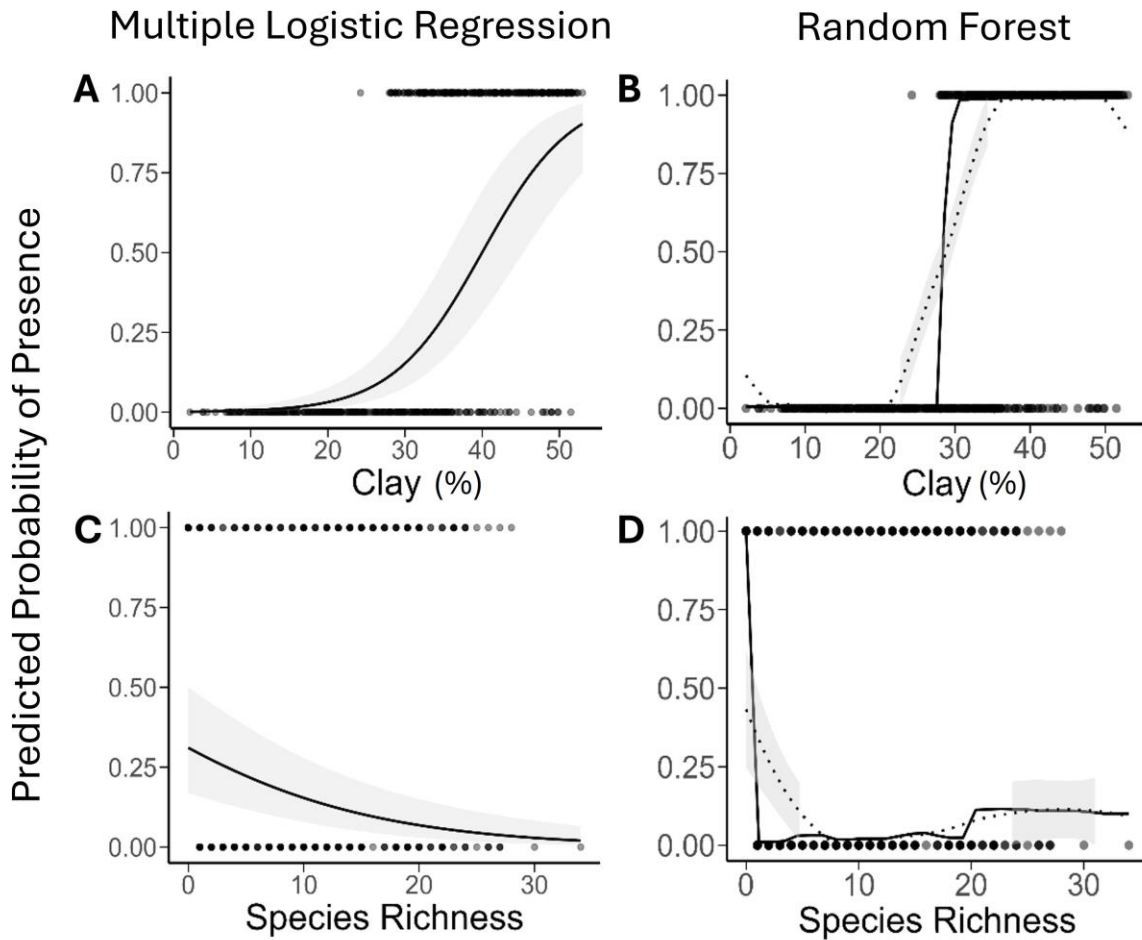


Figure 1.7. Predicted probability of presence in response to stream depth (m), number of dams, and baseflow index across the state of KS from (A), (C), (E) multiple logistic regression and (B), (D), (F) random forest ecological models of Topeka Shiner distributions in Kansas from 670 stream samples collected between 1947 and 2020. Predictor ecological variables include: (A), (B) *stream depth* (estimated in meters); (C), (D) *number of dams* (total in the mainstem), and (E), (F) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow) on the X axes. Response on the Y axes represents predicted probability of Topeka Shiner presence.

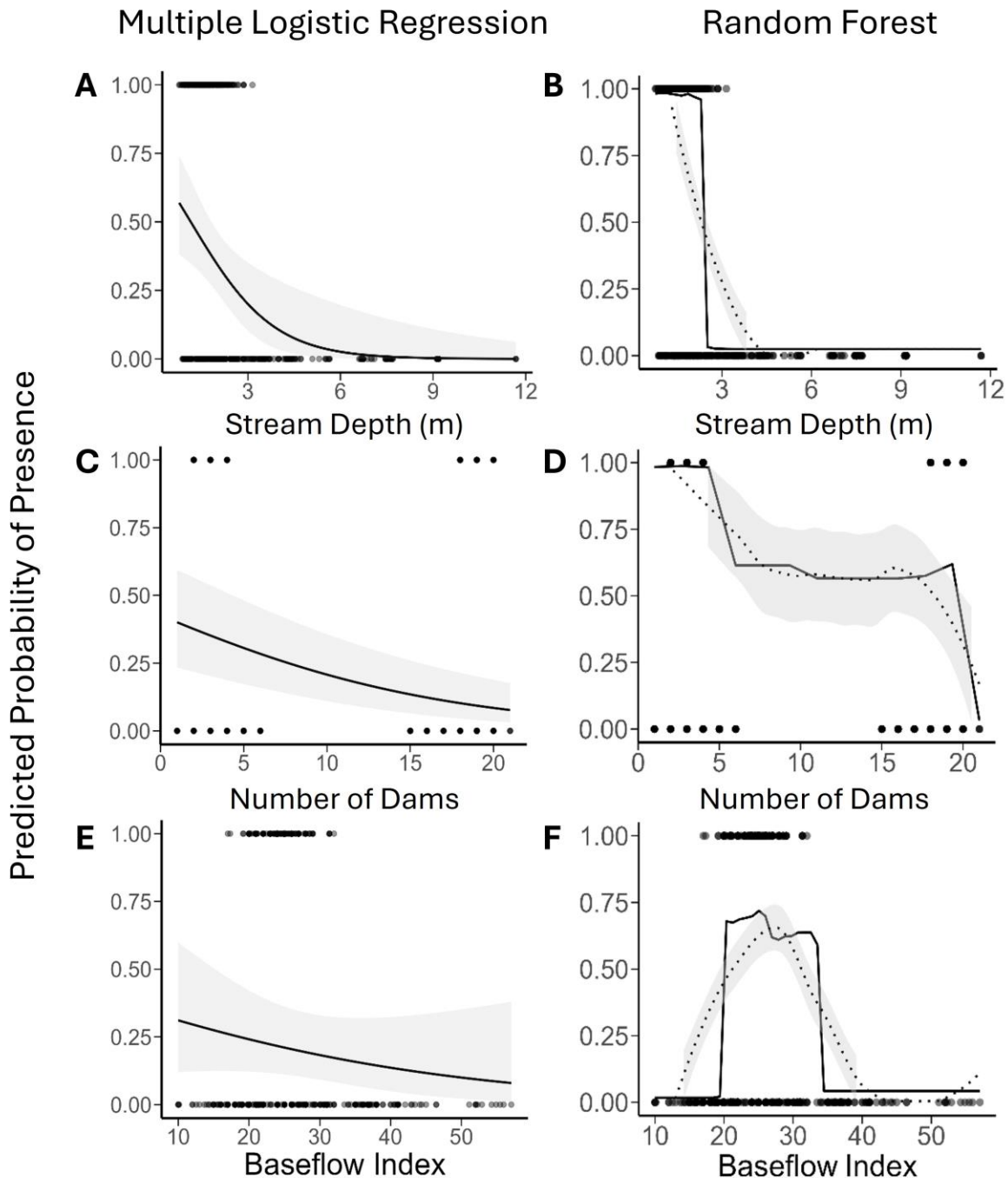
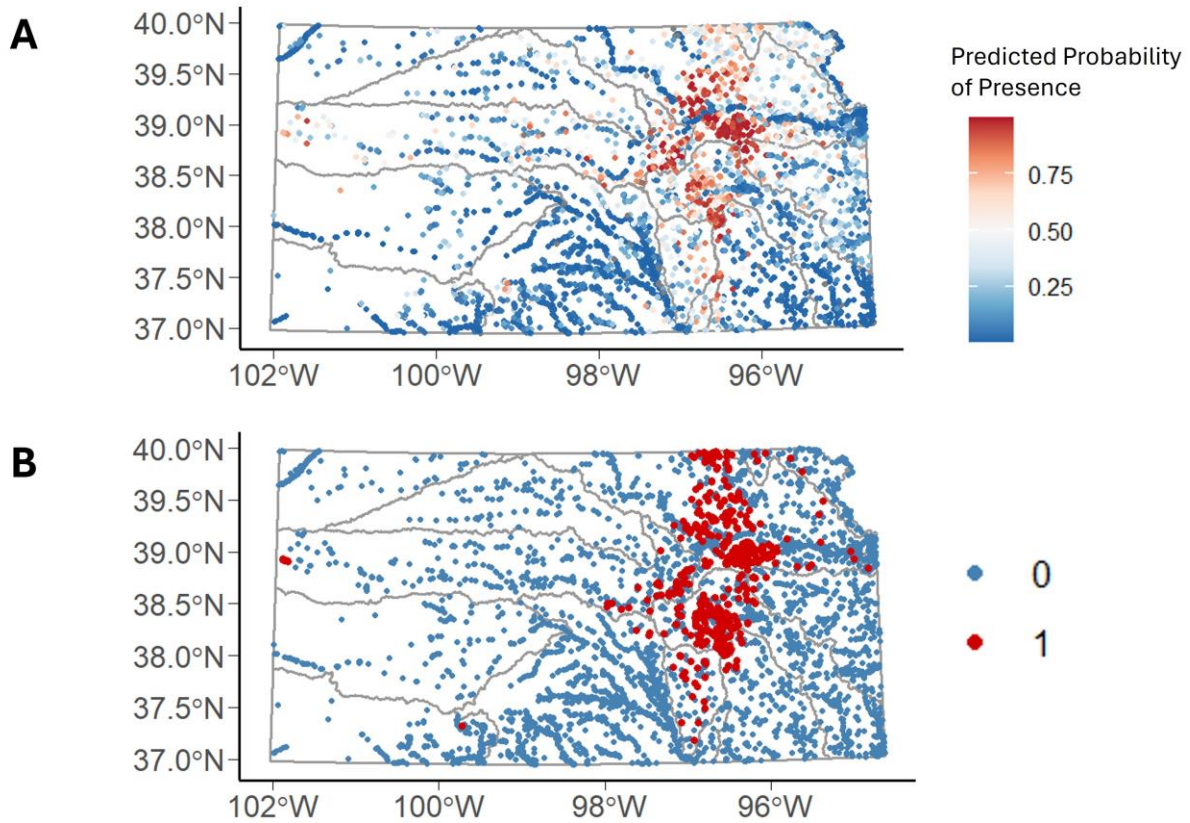


Figure 1.8. Map of predicted probability of Topeka Shiner presence in Kansas from 12,086 stream samples collected (81% between 2000-2021) as modeled by 11 ecological variables using (A) multiple logistic regression and (B) random forest. Comparisons were made between statistical approaches for sites above and below 50% probability of presence from multiple logistic regression and predicted presence (1) or absence (0) from random forest.



Chapter 2 - Applying and expanding a structured approach to the systematic statistical analysis and ecological interpretation of existing data to guide imperiled Great Plains fish conservation

Abstract

Data-based evidence supports conservation of native species at statewide and regional scales. However, analyses of statewide biodiversity monitoring data can generate highly uncertain trends for habitat associations of and threats to data-limited species. Our overarching goal is to analyze and interpret existing monitoring data to guide future data-driven activities, aid conservation planning, and support decision-making. To guide future research and management, we applied the Rode et al. framework to the Topeka Shiner (*Notropis Topeka*; $n = 335$ presence locations), using an existing statewide monitoring dataset ($N = 12,086$). In addition, we performed six connected iterations of this existing framework, each testing new data driven questions. These six iterations include (A) identifying agreement and disagreement in data and model trends, (B) testing nonlinear relationships, (C) refining ecological variable metrics and assessing changes in outcomes, (D) testing interactions between ecological variables, (E) incorporating random effects, and (F) comparing statistical models to identify consistently influential variables as needed. Comparisons across framework iterations advanced our ecological understanding of the drivers of Topeka Shiner distribution, regardless of subtle improvements in model accuracy (86 - 90%), precision (93 - 97%), and fit ($\Delta AIC_c \sim 50$). For example, consistency across all models in the significance of four ecological variables (*stream depth*, *clay*, *baseflow index*, and *predator species richness*) decreased ambiguity and increased confidence

in trends. Our analyses provide multiple examples of data-based conclusions, new questions of interest, related sampling strategies, and connections to agency applications. For example, future data collection activities could be critical to confirm an interactive effect between intermediate *baseflow index* (20-30%) and increasing *clay* (>30%) to inform habitat protections and reintroduction efforts based on hydrological regulating effects of baseflow. In summary, our application of the framework enables researcher-manager teams to identify novel and robust ecological trends from existing field data to support current and future data strategies for the conservation of Topeka Shiner.

Introduction

Importance of Data-Based Evidence for Effective Natural Resource Conservation.

The foundation of applied ecological research, conservation planning, and actionable management is often data-based evidence about drivers of natural and altered organismal patterns and related anthropogenic impacts (Salafsky et al., 2019). Biodiversity monitoring data provide an important source of ecological evidence for local, regional, and statewide status and trends of native and invasive species (Lindenmayer & Likens, 2010). Specifically, regular monitoring assessments of stream and river biotic communities are essential for quantifying native fish distribution and abundance, a type of data that helps to identify effective future conservation actions (Radinger et al., 2019). Our overarching goal is to systematically analyze and interpret existing monitoring data to better guide future data-driven activities, aid conservation planning, and support decision-making.

Challenges with Analyzing Existing Monitoring Data.

Monitoring datasets can require special consideration in data analysis and interpretation for several reasons. Statewide and regional monitoring programs collect organismal and environmental data across broad spatial and temporal scales representing a wide, natural range of abiotic and biotic conditions (e.g., KDWP, 2008; KDHE, 2020). As a strength, monitoring data for organisms of conservation concern provide a rare view of distributions and abundance at jurisdictional levels relevant to policy decisions (Lindenmayer et al., 2022). Nevertheless, difficulties can arise in analyzing and interpreting field monitoring data for which underlying biotic and abiotic conditions are highly variable and/or only partially constrained. First, an inevitable consequence of

providing these big picture trends is an increased level of unexplained variation that typically exceeds the magnitude of statistical error that is present in controlled and experimental datasets (Gitzen et al., 2012). Second, uncommon or imperiled species, which are often the focus of data-driven conservation, are present at fewer locations, have less data available, and often are associated with analyses that have lower statistical power (Olden & Jackson, 2002). Third, sampling designs of natural systems should consider the wide range of underlying abiotic heterogeneity that often occurs at statewide or regional scales (e.g., randomly stratified, incremental, or spatially-balanced sampling approaches; Steen et al., 2021). Fourth, field data sets often violate statistical assumptions (e.g., distributions, correlations), which can create highly biased and erroneous results (Olsen et al., 1997). Finally, because no sampling regime can sample everywhere all the time, even the best monitoring designs will have information gaps (Bowler et al., 2025). If unaddressed, these statistical challenges related to analyzing field data can increase ambiguity in results within and across studies, cause unjustified lack of significance, create spurious statistical patterns, and unnecessarily increase unassigned error. All these quantitative difficulties related to field data collected in monitoring surveys can prevent professionals, who are charged with aquatic and terrestrial conservation, from learning all they can from existing data.

Great Plains Stream Fish.

Improving analysis and interpretation of existing distribution data and associated threats can benefit the conservation of Great Plains native prairie stream fish. Prairie streams are a unique type of aquatic ecosystem, characterized by low gradients, naturally-occurring extreme flows, intermittent drying, turbid conditions, and a reduced canopy (Wishart, 2004). These aquatic systems, set in

grassland catchments, stretch from Texas to Montana (Matthews, 1988). In general, native fish species that inhabit Great Plains stream and river systems are becoming increasingly imperiled by increasing anthropogenic pressures that disrupt the naturally-occurring extreme flows and highly turbid water in which many of these small-bodied native fish have co-evolved (Dodds et al., 2004). Specifically, native prairie fish can be adversely affected by widespread predator sportfish stocking, agricultural impacts on water quantity and quality, land-use alterations that result from human development, and loss of connectivity caused by dam construction (e.g., Bouska & Paukert, 2010; Cross & Moss, 1987; Gido et al., 2010; Perkin et al., 2015). In Kansas, the prairie stream fish communities that inhabit more than 160,000 river kilometers of streams and rivers are a major conservation priority, because population declines have occurred in more than half of the 116 native fish species (Haslouer et al., 2005; Rohweder, 2022). Specifically, 32 native Kansas fish species are listed as species in need of conservation (SINC), 11 fish taxa are listed as threatened, and 5 species are considered endangered (T&E) (Rohweder, 2022). The current Kansas State Wildlife Action Plan (Rohweder, 2022) identifies that the distribution and abundance of native Kansas stream fish across the state's 12 major river basins are threatened by the same pressures that face prairie stream fish throughout the U.S. Great Plains. Underscoring the importance of conservation of these unique species is a need for data-driven insights to inform adaptive management efforts, especially for rare fish for which monitoring data are sparse.

Objectives.

Here we expand an existing adaptive management framework for analyzing fish monitoring data in Great Plains stream systems by addressing three objectives (Figure 2.1; Rode et al., 2025). Adaptive management frameworks can extract decision-relevant insights from limited available data (Williams, 2011) and integrate existing knowledge and data to create a strong foundation for ecological learning in ongoing conservation management (e.g. Bopp et al., 2025). Rode et al. (2025) developed a 12-step framework that addressed the conservation question of “*why might a specific native stream fish taxon be present or absent at specific locations?*” For our first objective, we apply the original Rode et al. (2025) framework to an imperiled fish species, the Topeka Shiner (*Notropis topeka*), using a statewide Kansas monitoring database. As a second objective, we applied the 12-step framework in six additional connected iterations to test new and increasingly complex questions to hone ecological insights and support more specific guidance for future data collection, monitoring analyses, and restoration options (Figure 2.1). These additional iterations included (A) identifying agreement and disagreement in the systematic review of data distributions and model trends, (B) testing for and including nonlinear relationships in statistical models as appropriate, (C) refining ecological variable metrics and assessing changes in outcomes, (D) testing for and including interactions between ecological variables in the statistical model as appropriate, (E) incorporating random effects to the statistical model as appropriate, and (F) comparing individual statistical models to identify consistently influential variables. For our third objective, in the discussion, we integrate knowledge across all iterations of the 12-step Topeka Shiner framework to generate ecological insights, inform specific conservation actions, and provide recommendations for future data collection and analyses (Figure 2.1). The novelty of this approach resides in its application to a rare fish taxon with limited monitoring

data, which demonstrates utility of the Rode et al. (2025) framework with implications for conservation monitoring of other rare taxonomic groups.

Methods

Overview of the 12 Steps in the Rode et al. (2025) Framework.

The Rode et al. (2025) adaptive management framework (hereafter, “the framework”) couples long-term monitoring data to a 12-step loop that can progressively accumulate ecological insights through repeated, connected cycles (Figure 2.1). In each 12-step framework iteration (cycle or loop), researcher-manager teams prioritize a single, relevant overarching question for the entire cycle (Step 1). A specific taxon and relevant spatiotemporal scale (Step 2) are matched to the overarching question (Step 1). While Steps 1-2 establish the focus of one framework iteration to a single overarching question, taxon, and scale, the 12-step framework cycle can be repeated multiple times to address additional overarching questions, taxa, and/or scales. Next, relevant literature is reviewed and analyzed through a Weight of Evidence (*WoE*) procedure (Step 3), and teams methodically select fish-relevant ecological variables from the literature (Step 4). Once existing habitat and impact data for sites with fish occurrence data are compiled (Step 5), teams clean and check habitat and fish data (Step 6). Then, a series of targeted questions are defined (Step 7) related to the overarching question (Step 1) and are next addressed by systematically running a standard set of data visualizations and statistical analyses in specific sub-steps (Sub-steps 8A-F). Outputs from all previous steps (especially Steps 3 and 8) are integrated through a qualitative

WoE procedure (Step 9), to identify emerging questions (Step 10), propose options for realistic field sampling designs to guide future efforts (Step 11), and link data, analyses, and plans to subsequent iterations of the framework (Step 12).

Applying the Framework to a Rare Taxon (Objective 1).

Steps 1-2. Overarching Question, Taxon, Scale. Here, our overarching question was “Using fish monitoring and existing GIS data, what environmental and biotic variables are linked to the site-specific probability of a Topeka Shiner being present or absent across Kansas, a US state?” We selected Topeka Shiner, because this federally-endangered and Kansas-threatened species is a high priority in management planning (KDWP, 2022; USFWS, 1998). For our Kansas statewide scale, we used 12,086 fish presence-absence samples for 10,624 unique stream sites from standardized monitoring, historical specimens, and collection permits (KDWP, 2024). Eighty-one percent of these collections were made between 2000 – 2021.

Step 3 - Literature Review: In a Web of Science literature review, we used the keywords “*Notropis topeka*” (or “Topeka Shiner”) and “Kansas” to compile original studies and review articles. While we included “Kansas” as a search term, relevant peer-reviewed journal articles could include studies undertaken across Topeka Shiner’s native range including sites outside of Kansas. To define expected trends with ecological variables, we calculated the percentage of these peer-reviewed papers that reported significant trends for each variable considered. Articles that reported qualitative, observational, untested trends and non-significant effects were grouped and compared to significant trends.

Steps 4 –5 Choose and Acquire Environmental Data: From five GIS databases and existing monitoring data (Table 2.1), we obtained variable metrics for the following eight ecological categories: *stream size*, *substrate*, *stream flow*, *biota*, *land use*, *gradient*, *curvature*, and *dams*. We carefully assessed how the ecological variables could influence fish survival and growth then defined specific, measurable metrics to characterize these variables (Table 2.1). We prioritized recent (2000 – 2021) GIS data corresponding to years in which most stream fish were collected (81%) to calculate environmental variable values for all sites. We extracted GIS data for the same latitudes and longitudes as fish samples.

Step 6 – Data Cleaning. We performed quality control measures to clean GIS data and stream site samples by removing entries with missing values, checking and correcting location data, and assessing influence of extreme values. We transformed the abundance of all fish species including Topeka Shiner to presence-absence at the site level, to account for effects of collection gear and year variation on abundance estimates.

Step 7 – Specific Questions. As in the original proof-of-concept application of the framework (Rode et al., 2025), we chose six specific questions that correspond to subsequent quantitative procedures in Sub-steps 8A-F. These specific questions included: (A) Where is the taxon present and absent?; (B) What is the range of environmental conditions sampled?; (C) How do individual environmental variable values differ between fish presence and absence sites?; (D) What are the best metric and scale option for each ecological variable?; (E) What is the quality of the best statistical models created (e.g., accuracy, precision)?; and (F) How are ecological variables associated with fish presence and absence in the best statistical model(s)?

Step 8 – Quantitative Loop. In sub-steps 8A-F, we match a specific visualization or analytical procedure to each specific question listed above (Step 7A-F). Presence-absence maps (Sub-step 8A; specific question 7A) were developed in ArcGIS Pro (Esri Inc., 2024). All other analytical procedures were performed in program R (R Core Team, 2025). For Sub-step 8B, we created histograms (*ggplot2*; Wickham, 2016), while bar plots, spatial maps, kernel density estimator (KDE) ridge plots, boxplots and simple logistic regression probability plots (*sf*, *stats*, *ggplot2* packages; Pebesma, 2018; R Core Team, 2025; Wickham, 2016) were created for Sub-step 8C. For Sub-step 8D, we selected a single metric within each ecological category by prioritizing the metric with the most consistent effects on Topeka Shiner presence, the fewest missing values, and minimal zero-skewness.

For Sub-steps 8E-F, we balanced our Topeka Shiner presence-absence data by randomly selecting an equal number of absence sites to pair with Topeka Shiner presence sites (Steen et al., 2021). A small number of sites ($n = 10$ of 670 observations) had repeated sampling efforts on separate dates, which was insufficient to assess change in Topeka Shiner distributions and ecological relationship through time. Given the very small proportion of affected samples, we retained all observations. We excluded highly correlated metrics (Pearson's correlation threshold ≥ 0.70) from the same model and ran separate models for correlated variables (reported in Appendix B Table 3-Appendix B Table 4). With the *performance* package, we assessed multicollinearity using variable inflation factors (VIF) > 10 (Kutner, 2005; Lüdtke et al., 2021). We fit multiple logistic regression (MLR) models with the original and standardized data (mean = 0; standard deviation = 1), a logit-link function, and a Laplace maximum likelihood estimator from the *glmmTMB* package (Hosmer et al., 2013; McGillicuddy et al., 2025). Top models were identified as those for which $\Delta AIC_c \leq 2$ (*AICcmodavg* package; Arnold, 2010; Mazerolle, 2023; Sutherland et al., 2023). We evaluated the performance and sensitivity of top

model(s) with 10-fold cross validation and the area under receiving operator characteristic curve (*caret*, *pROC* packages; Kuhn & Max, 2008; Robin et al., 2011). For significant variables in the top model ($p \leq 0.05$; 95% CI excludes zero), we evaluated the direction, size, and odds-ratio (OR) of variable-specific standardized beta coefficients. With probability plots, we visualized the marginal effects of individual environmental variables by holding all other variables in the ecological model at the mean (*stats*, *ggplot2* packages; R Core Team, 2025; Wickham, 2016). In the discussion, we integrated evidence from the literature review, visualizations, and analyses to determine the role of each variable using *WoE* (Step 9), as well as identify emerging questions, potential next steps, and links to continued framework iterations (Steps 10-12). Summaries of this first and subsequent iterations of the framework to address connected emerging questions are provided in (Table 2.2). For Objectives 2A-F, we followed and built upon the same procedures for MLR modeling as described in Objective 1.

Framework Iterations (Objective 2; Figure 1).

Compare Data Visualizations Systematically and Critically (Objective 2A).

For the second framework iteration, we asked, “How do data visualizations and modeled trends agree or disagree on the influences of environmental variables on Topeka Shiner distributions?” We critically examined six data visualizations for each environmental variable and predicted that informative similarities and differences would be identified across plots. Specifically, we identified consistent and inconsistent trends across histograms, boxplots, ridge plots, habitat maps, and probability plots to aid interpretation of statistical models. We expected broad agreement between patterns observed for environmental variables previously

associated with Topeka Shiner habitat, while recognizing that differences could arise due to weaker correlative relationships, insufficient variation of data, existence of nonlinear trends, and interactions among environmental variables.

Exploring Nonlinear Variable Parameterization (Objective 2B).

For the third framework iteration, we addressed the emerging question, “Does *baseflow index* influence Topeka Shiner distributions as a nonlinear variable?” We predicted based on inconsistent data trends identified in Objective 2A that *baseflow index* would exhibit a nonlinear relationship with Topeka Shiner presence, with occurrence highest at intermediate values of *baseflow index*. Therefore, we explored nonlinear trends by evaluating three parameterizations of the variable *baseflow index*. Candidate simple logistic regression models that explored how these linear and nonlinear parameterizations of *baseflow index* affected Topeka Shiner presence included: (1) one linear parameter, (2) one squared parameter, and (3) two quadratic parameters (linear and squared terms together). Then, the most explanatory parameterization ($\Delta AIC_c < 2$) was reintroduced to the full multiple logistic regression model. As needed for polynomial variables, we calculated inflection points (e.g. peaks) with original data and maximized instantaneous slopes with standardized data (Strang & Herman, 2016a, 2016b). Due to including variables twice in the same model to accommodate a polynomial term, we expected inherently high (>10) VIF with original data (Kutner, 2005).

Refine Specific Metrics within an Ecological variable (Objective 2C).

For the fourth framework iteration, we asked, “What groups of species provide more specific insights about the negative relationship with high species richness?” We predicted that subsets of the fish community defined by biotic interactions and species traits would have stronger associations with Topeka Shiner distributions than diversity of the entire fish community. By replacing the biotic metric used in our previous iterations, *total species richness*, we explored three biotic metrics (*predator species richness*, *competitor species richness*, *facilitator species richness*) to provide more actionable information on functional relationships within fish communities. The biotic metrics were based on guild classifications developed from a national database (FishTraits database; Frimpong & Angermeier, 2009), recommendations by a Kansas fish expert (J. Hofmeier, personal communication, 2024-2025), and supporting Kansas literature (e.g., Eberle et al., 2014). We assigned fish species to the three trait groups based on (1) trophic ecology, (2) body size, and (3) reproductive ecology using characteristics described for adults (Appendix B Table 1). First, species with piscivorous diets and body sizes larger than Topeka Shiner (>7.6 cm) were placed into a “potential predator guild.” Second, we assigned species with invertivorous diets to a “potential competitor guild.” Finally, fish attributed with nest building and egg guarding behaviors (including lithophiles) comprised a “potential facilitator guild.”

Explore Interactions (Objective 2D).

For the fifth framework iteration, we addressed the emerging question, “Do multiple environmental variables interact to influence Topeka Shiner distributions simultaneously?” We tested the significance of two and three-way interactions among the significant variables ($p \leq 0.05$) identified in Objective 1 and prioritized interactions with ecological justifications. For example, we

predicted that the influence of hydrological conditions would vary with other physical stream habitat characteristics (Rosenfeld, 2017), such that combinations of environmental variables could produce increases, decreases, or peaks in predicted probability of Topeka Shiner presence beyond the effects of individual variables. To simplify the complexity of Objectives 2D-E , we only included significant variables in the final model iterations (Bolker et al., 2009). For significant interactions, we visualized marginal effects to understand how the response of one explanatory variable on the predicted probability of Topeka Shiner presence varied across levels of the second explanatory variable (*interactions package*; Long, 2024).

Account for Random Effects (Objective 2E).

For the sixth framework iteration, we asked “Are trends with environmental variables stable despite the distribution of our limited data?” With regional concentration of Topeka Shiner presences, we predicted that accounting for regional variation would improve estimates of environmental relationships at the statewide level. We utilized proportional bar plots to assess the relative distributions of samples collected in our balanced dataset across the 12 major KS river basins (*ggplot2 package*; Wickham, 2016). To detect spurious and spatially correlated relationships from previously described trends across the state, we incorporated a random intercept in the reduced ecological model through grouping sites by major river basins. Furthermore, we evaluated model predictions by applying the mixed model estimates to the full monitoring dataset, which included samples not used in training (*stats package*; R Core Team, 2025). Then, we assessed predictions spatially to compare mapped model estimates with the map of original fish observations (*ggplot2 package*; Wickham, 2016).

Compare Models (Objective 2F).

For the seventh framework iteration, we addressed the emerging question, “What do we learn from comparing results from each model in Objectives 1-2E?” Similar to Objective 2A, we predicted that comparisons among models would reveal informative similarities and differences in ecological inference. We expected agreement among models for the most influential environmental variables, while disagreements would reflect the effects of different modeling assumptions or weaker ecological relationships given the current dataset. We further predicted that sequentially incorporating ecological complexity would refine inference by revealing whether additional model structures improved explanation and prediction of Topeka Shiner distributions. Specifically, we identified consistent and conflicting trends and changes in information criteria, model accuracy, and relative strength of relationships from framework and MLR model iterations. Rankings for relative strength (or influence) of ecological variables on predicted probability of Topeka Shiner presence were assigned using the size of standardized coefficients (e.g., 1st = highest absolute magnitude), or the combined instantaneous slope for polynomial variables.

Results

Objective 1 – Application of the Framework to a New Taxon (Steps, 3, 7, 8A, 8C, 8E-F, 9).

Step 3 - Literature WoE. Inconsistency in the current literature about how variables affect Topeka Shiner makes establishing generality challenging. Conclusions from 35 peer-reviewed studies did not consistently agree on whether and how variables in

different ecological categories affected Topeka Shiner (Figure 2.2; Appendix B Table 2). The eight studies that measured *stream length, order, or depth* reported divergent results about whether *system* size was positively (25%; e.g. Galic et al., 2023), negatively (50%; e.g. Winston, 2002) or not significantly related (25%; e.g. Fischer et al., 2018) to Topeka Shiner presence (Figure 2.2A). The five studies that measured *substrate* as *bedrock geology, sand, gravel, or fine-to-coarse particles* provided mixed views on whether any of these *substrate* metrics significantly affected Topeka Shiner (40% positive sand and alluvium, e.g. Witte et al., 2009; 60% no significant effect, e.g. Gerken & Paukert, 2013; Figure 2.2B). Six studies differed in conclusions about Topeka Shiner's relationship with in-channel *surface flow* (33% positive; Wall et al., 2004; 66% no significant effect; Adams et al., 2000; Bouska & Paukert, 2010; Figure 2.2C) and *groundwater flow* (33% positive; Wall et al., 2004; 66% no significant effect; Bayless et al., 2003; Wall & Berry, 2004; Figure 2.2C). Of seven articles, researchers disagreed about the importance of various *biotic influences* including community diversity (66% positive; Gerken & Paukert, 2013; Simpson et al., 2019; 33% no significant effect; Schrank et al., 2001), predators (20% negative; Schrank et al., 2001; 80% no significant effect; e.g., Winston, 2002), competitors (17% positive; Simpson et al., 2019; 17% negative; Simpson et al., 2021; 66% no significant effect; e.g., Gerken & Paukert, 2013), and nest facilitators (100% positive; $n=1$; Simpson et al., 2021; Figure 2.2D). Related to *land use*, three studies diverged in assessments of how *pasture* and *row crop agriculture* influence Topeka Shiner (33% positive; 66% no significant effect), and one study queried the role of urbanized development (no significant effect; Gerken & Paukert, 2013; Figure 2.2E). Two empirical studies deviated in judgement about the impact of *dams* on Topeka Shiner (50% negative; Schrank et al., 2001; 50% no significant effect; Gerken & Paukert, 2013; Figure 2.2F).

Steps 7, 8A. In Kansas, Topeka Shiner was present in 341 stream monitoring samples from 335 unique locations (Figure 2.3). Presences occurred in the Lower Kansas-Republican River, Neosho, and Smoky Hill-Saline River basins, primarily in the Flint Hills ecoregion of Northeast Kansas. A single, more isolated population occurred near the western Kansas border.

Steps 7, 8C. The relationships with Topeka Shiner presence differed across ecological variables. In exploring the raw data, Topeka Shiner presence was inversely related to *stream depth* with more presences at sites with low (< 3m) *stream depths* and more absences in deeper streams (Figure 2.4A). More Topeka Shiner presences occurred in stream sites with higher *clay* compositions (> 30%) (Figure 2.4B). Most Topeka Shiner presences occurred at sites with a *baseflow index* between 20-30% whereas this imperiled fish taxon was often absent at *baseflow index* below 10 and above 30% (Figure 2.4C). Topeka Shiner was most often present in streams with lower *species richness* (Figure 2.4D). Fish presences often occurred at higher values of *grassland*, *gradient* (~0.005), *curvature* (>28) (Figure 2.4E, F, G). The total number of sites and Topeka Shiner were collected more often in streams with lower *development*, *cropland*, *pastureland*, and *number of dams* (Figure 2.4H-K).

Steps 7, 8E, 8F. Our top MLR model had 88% accuracy and 93% precision, without high multicollinearity ($VIF < 10$). This top model identified four ecological variables that were significantly related to Topeka Shiner presence: *stream depth*, *clay*, *species richness*, and *number of dams* (Table 2.3). Using standardized data, our top model predicted that Topeka Shiner presence was associated with decreasing *stream depth* ($\beta = -1.86$), with increasing *clay* ($\beta = 2.07$), lower *species richness* ($\beta = -0.69$), and a reduced *number of dams* ($\beta = -0.79$).

Step 9 – WoE. From the literature (Step 3), we identified that *stream size, substrate, flow, biota, land use, and dams* were important variables in some Topeka Shiner studies. In our data assessments, agreement existed among visualization plots (Sub-step 8C) and our MLR model (Sub-steps 8E-F) about the importance of *stream depth, clay, species richness, and number of dams*. Combined literature, visualization, and statistical model data provided a strong foundation for making practical decisions on next steps and future conservation actions related to Topeka Shiner (Steps 10-12; see discussion).

Objective 2A. Compare Data Visualizations Systematically and Critically.

Both consistent (e.g., *stream depth*) and inconsistent (e.g., *baseflow index*) trends emerged from *WoE* integration of data visualizations (Sub-steps 8B). For both variables, histograms showed that a range of sites were sampled (

Figure 2.5A, Figure 2.6A), and that no obvious regional patterns existed (

Figure 2.5D, Figure 2.6D). For *stream depth*, boxplots (

Figure 2.5B) and KDE ridge plots (

Figure 2.5C) agreed that Topeka Shiner was most often present at shallow *stream depths* (Sub-Step 8B), and probability plots from single and MLR further confirmed negative associations between Topeka Shiner and *stream depth* (

Figure 2.5E-F). However, for *baseflow index*, boxplots and KDE ridge plots showed that Topeka Shiner most often occurred at an intermediate level of *baseflow index* (Figure 2.6B-C), but probability plots from single and MLRs indicated negative associations between Topeka Shiner presence with *baseflow index* (Figure 2.6E-F).

Objective 2B. Explore Nonlinear Variable Parameterization.

In assessing whether a non-linear function better describes the relationship between *baseflow index* and probability of Topeka Shiner presence, we found that individual parameters (*baseflow index* or *baseflow index*² alone) did not describe data trends well ($\Delta AIC_c > 2$; Figure 2.7A, B). However, using a quadratic function (*baseflow index* and *baseflow index*² together) identified a significant downward parabolic relationship (peak-shaped curve) between *baseflow index* and Topeka Shiner probability of presence (Figure 2.7C). In adding a quadratic parameter for *baseflow index* to our MLR with the same 11 ecological variables described earlier, *stream depth*, *clay*, *species richness*, and *number of dams* continued to significantly affect Topeka Shiner probability of presence. In addition, a significant relationship emerged between *baseflow index*² and the probability of Topeka Shiner presence, showing the importance of the combined quadratic effect. Furthermore, the inverted parabola showed that predicted probability of Topeka Shiner presence peaked at 24.4 *baseflow index* (Figure 2.8), and the strongest response to one standardized increase in *baseflow index* was 6.63-fold increase in predicted probability of Topeka Shiner presence (combined slope at minimum *baseflow index*; individual estimates in Table 2.4).

Objective 2C. Refine Metrics within an Ecological variable.

In examining relationships between specific subsets of the fish community and probability of Topeka Shiner presence, we found small, but informative, ecological and statistical differences in the effects of *predator species richness* and *competitor species richness* compared with the original biotic variable, *total species richness* (Figure 2.9). The relationship between *facilitator species richness* and the probability of Topeka Shiner presence was weaker and more variable (Figure 2.9). For all four biotic metrics, communities with lower *species richness* were sampled most often (Figure 2.9A-D). KDE ridge plots showed clear decreasing trends, in which both Topeka Shiner presences and absences were more common with lower *species richness* (Figure 2.9E-H). On average, stream communities at sites with higher *total species richness* (Figure 2.9I), *predator species richness* (Figure 2.9J), and *competitor species richness* (Figure 2.9K) included Topeka Shiner less often than communities with fewer species. *Facilitator species richness* did not show a clear effect on Topeka Shiner presence or absence, which may be in part due to nest-building species occurring at relatively few sites overall (Figure 2.9L). Probability plots from single logistic regression also showed significant inverse relationships between Topeka Shiner presence and *total species richness* (Figure 2.9M), *predator species richness* (Figure 2.9N), and *competitor species richness* (Figure 2.9O). Among the subsets of the fish community, only *predator species richness* demonstrated stronger negative associations and higher explanatory power ($\Delta AIC_c < 2$) than *total species richness*. When introduced to the MLR model, *predator species richness* (Table 2.5), like the previous model that contained *total species richness* (Objective 2B, Table 2.5), showed significant relationships between Topeka Shiner presence and *stream depth*, *clay*, *quadratic baseflow index*, and *number of dams* (Table 2.5; Figure 2.10A, B, C, D, K). As a new insight, when the significant *predator*

species richness replaced *total species richness* as a biotic variable, *grassland* became a significant environmental variable and had an increasing relationship with the predicted probability of Topeka Shiner presence (Table 2.5; Figure 2.10E).

Objective 2D. Explore Interactions.

When examining two- and three-way interactions among all variables for multiple combinations of ecological conditions that may support Topeka Shiner persistence, only an interaction between *baseflow index* and *clay* (Table 2.6) was significantly related to predicted probability of Topeka Shiner presence. As with previous models (Objective 2C; Table 2.5), the effects of *stream depth*, *predator species richness*, and *number of dams* continued to be significant (Table 2.6). The significant interaction between *baseflow index* and *clay* predicts that either variable may compensate for conditions associated with lower probability of Topeka Shiner presence. For the greatest probability of Topeka Shiner presence (75%; Figure 2.11), sites conditions included approximately 50% *clay* and 20-40 *baseflow index*. However, even with lower *clay* composition, the probability of Topeka Shiner presence exceeded 50% in the peak range of *baseflow index* (20-30 %), (Figure 2.11A). Furthermore, increases in *clay* composition boosted the predicted probability of Topeka Shiner presence beyond 50% despite lower *baseflow index* (Figure 2.11B).

Objective 2E. Account for Random Effects.

In considering whether distributional biases may have influenced trends with probability of Topeka Shiner presence, we found that most of the samples and all Topeka Shiner presences occurred in three of the 12 major Kansas river basins (Figure 2.12A-B). With random effects to offset the weight of sites in the more prevalent basins in our MLR model (Objective 2E, Table

2.7), *stream depth*, *clay*, *quadratic baseflow index*, *predator species richness*, and the interaction between *clay* and *baseflow index* (Objectives 1, 2B, 2C, 2D; Table 2.3-Table 2.7) continued to show significant relationships with predicted probability of Topeka Shiner presence. The inclusion of random effects confirmed that the trends we detected were common across major river basins. However, the *number of dams* no longer contributed significantly to statewide predictions of Topeka Shiner presence in the random effects model (Table 2.7), because the random effect absorbed variability previously attributed to this variable. From random effect model predictions, high (> 50%) probabilities of Topeka Shiner presence concentrated in the Kansas-Lower Republican Basin and resembled the observed distribution of presences (Figure 2.3, Figure 2.13). Low (<50%) probabilities of Topeka Shiner presence spread across most of the remainder of the state, including the Upper Republican, Upper and Lower Arkansas, and the Missouri River Basins.

Objective 2F. Compare Models.

All five of our model iterations [(1) original framework (Objective 1); (2) non-linearity (Objective 2B); (3) refinement in metrics (Objective 2C); (4) interactions among variables (Objective 2D); (5) random effects (Objective 2E)] showed consistency in identifying variables that significantly affected Topeka Shiner presence (Table 2.8). For all models, *stream depth* and *species richness (total or predator)* always had significant negative effects on probability of Topeka Shiner presence, and *clay* had a significant positive effect (Table 2.8). When *baseflow*² was included in models (Objectives 2B-2E; Table 2.8), *quadratic baseflow index* was consistently significant. Similarly, when the interaction between *baseflow index* and *clay* was tested (Objectives 2D, 2E; Table 2.8), this effect was consistently significant. Although accuracy (88, 88, 89, 86, 90%), precision (93, 94, 93, 93, 97%), and sensitivity (85, 84, 85, 83, 85%) were

high in all models, AIC_c decreased as the model sophistication increased (Table 2.8; Objective 1 → Objective 2B → Objective 2C → Objective 2E, but not for Objective 2D).

Some informative differences emerged across models. First, in many models, *numbers of dams* significantly decreased the predicted probability of Topeka Shiner presence (Table 2.8; Objectives 1, 2B-2D), but the lack of significance of the *numbers of dams* variable in the random effects model showed that this variable effect could perhaps be explained by differences across river basins (Table 2.8; Objective 2E). Second, a positive, significant relationship with *grassland* emerged only when *predator species richness* was substituted for *total species richness* (Objective 2C, Table 2.8), suggesting that the broader biotic metric may have masked a historic land use effect.

Comparing model coefficients (β) within and across models provided additional useful information for evaluating the strength of relationships. Most standardized regression coefficients changed minimally. In all models, *clay* and *stream depth* had consistently high standardized regression coefficients (Table 2.8). In contrast, changes in the *number of dams* and *species richness* exhibited relatively lower influences on predicted probability of Topeka Shiner presence (Table 2.8; 3rd to 6th). However, as model sophistication increased, the relative influence of *clay* decreased (Table 2.8; 1st to 4th), and the relative influence of *stream depth* and *baseflow index* increased (Table 2.8; 1st and 2nd). Furthermore, predicted probability of Topeka Shiner presence exhibited a stronger response to the interaction between *clay* and *baseflow index* than the main effect of *clay* alone in the final model (Table 2.8; β : $|-1.78| > 0.85$).

Discussion

Overview.

Our systematic application of the Rode et al. (2025) framework to existing state-wide monitoring data for an imperiled species, Topeka Shiner, has provided new quantitative and ecological insights that make statistical analyses of field data more useful, usable, and defensible. As a result of the data analyses presented here, three broad take-home messages emerged that are discussed in detail below. First, we confirmed that useful insights could be gained from applying the original 12-step framework to existing field data on the federally endangered and state-threatened Topeka Shiner in Kansas. Second, we refined the representation of ecological relationships by adapting existing framework procedures to iteratively test new data questions. Third, we provide guidance on *where* and *how* targeted sampling can address knowledge gaps through our systematic analysis. These three take-home messages can help conservation professionals make evidence-based decisions on how to allocate limited time and resources.

Take-Home 1: Applicability of the Original Framework to Topeka Shiner Data (Objective 1).

By applying the original 12-step framework to Topeka Shiner distributional data, useful insights emerged about previously described drivers (*stream depth*, *species richness*, *number of dams*), as well as new (*clay*) environmental correlates. First, our results confirmed the previously reported Kansas pattern of Topeka Shiner presence decreasing as *stream depth* increases (Mammoliti 2014). The Topeka Shiner literature identifies varied influences of system size on this imperiled fish, and our general categorization of Topeka Shiner as a small stream system fish inhabiting shallower sites is consistent with existing literature showing that Topeka Shiner

inhabits shallower depths (Stark et al., 2002), lower order streams (Winston, 2002) in smaller drainage systems (Wahl et al., 2024).

Second, our data analyses confirmed that the probability of Topeka Shiner presence is greater in Kansas systems with lower *species richness*. The peer-reviewed literature for Topeka Shiner in Kansas examined varied influences of the biotic community, which have varied effects on Topeka Shiner. Our finding about *species richness* is consistent with evidence directly linking Topeka Shiner presence to low stream biotic diversity (Mammoliti, 2014; Schrank et al., 2001) and indirectly linking less diverse communities to smaller stream systems (Fletcher Jr. et al., 2025; Vannote et al., 1980). However, other studies suggest that Topeka Shiner can coexist with more species in deeper habitats within streams such as pools or oxbows (Galic et al., 2023; Schrank et al., 2001; Simpson et al., 2019). This inconsistent response to system size and biota merits further investigation, because our predicted global trends for Topeka Shiner response to whole stream segment measurements may differ from fish responses to ecological variables measured at local habitat scales (Fausch et al., 2002; Steel et al., 2010).

Third, our application of the framework analyses to Topeka Shiner also identified *number of dams* as a significant adverse impact. As suggested by general prairie stream literature (Dodds et al., 2004; Ward et al., 1983), habitat and flow alteration from dam construction can be detrimental to small-bodied native stream fish communities. We found that Topeka Shiner presence was associated with fewer dams, a trend that is generally consistent with expectations from our literature examining smaller impoundments (Gerken & Paukert, 2013; Schrank et al., 2001). Although, this association diminishes in our later models (Objective 2), and the substantial overlap in dam counts between sites occupied and unoccupied by Topeka Shiner

could suggest that other characteristics of anthropogenic barriers (e.g., size, function, location) as opposed to number may better explain Topeka Shiner distributions in Kansas.

Fourth, we identified *clay* as a novel variable that is influential in predicting Topeka Shiner presence. In our original model (Objective 1), *clay* soil was statistically significant, had a large parameter estimate ($\beta = 2.07$), and probability of Topeka Shiner presence was highest among stream sites with greater than 30% surface horizon *clay* composition in the surrounding landscape. Unlike most previous studies of Topeka Shiner which focus on in-channel substrate [i.e., gravel substrate in the stream bed (Witte et al., 2009); but see (Schrack et al., 2001; Wall et al., 2004)], we used a measurement of substrate as soil composition at broad spatial extents (here, stream catchments). While individual fish may or may not select for instream *clay* substrate, we found consistent indicators that systems with *clay* in the surface horizon geology exerted strong influences on Topeka Shiner distributions. These geological relationships at the larger landscape and watershed scales present a useful new avenue for understanding Topeka Shiner distributions with an important yet under-explored ecological relationship.

Take-Home 2: Multiple Connected Framework Iterations can Enhance Ecological Insights (Objective 2).

With demonstrating how multiple connected iterations operate in the Rode et al. (2025) framework, we tested data-driven questions that required the addition of more complex features to our existing models (Objective 2). Starting with simple models has advantages in statistics (e.g. Bolker, 2008), fisheries monitoring (e.g. Dudley et al., 2022), and adaptive management of environmental flows (e.g. Webb et al., 2017). Specifically, more complex models can be difficult to interpret ecologically (Zuur et al., 2009) or provide weaker evidence for traditional hypothesis

testing (Ranganathan et al., 2016). Nevertheless, given the challenges of field data, running and comparing sequentially more complex models with ecological justifications can reveal useful trends, explain ambiguity, and guide future data collections.

By examining and comparing multiple visualizations (Objective 2A), we identified inconsistencies that can help in variable metric selection, suggest appropriate parametrization, and diagnose assumption violations. Our systematic and detailed comparison of data plots can be useful in alerting analysts to potential discrepancies prior to analysis and interpretation. Similar trends in data plots (*stream depth*) can provide confidence in data interpretation initially, during analysis, and in final synthesis. Differences in plots of the same data (*baseflow index*) can alert the analyst that complexity exists that needs further examination. As an example, in our initial analysis, data distributions and preliminary trends for *baseflow index* disagreed.

To address inconsistency across data plots, we reparametrized a nonlinear, quadratic relationship between *baseflow index* and Topeka Shiner presence (Objective 2B). The importance of groundwater proxies such as a nonlinear *baseflow index* would have been previously missed in our analysis and other models (e.g. Bayless et al., 2003) if we had only examined linear relationships. However, identifying the peak-shaped relationship between *baseflow index* and Topeka Shiner presence was critically important because this parabolic statistical relationship can reflect a positive, neutral, or negative ecological effect on Topeka Shiner presence depending on levels of *baseflow index* in any given system. For example, the peak in predicted probability of Topeka Shiner presence at 24.4 *baseflow index* represents a potential habitat threshold at which further increases in *baseflow index* may result in lower likelihood of Topeka Shiner presence in streams. In summary, by critically comparing data

visualizations and adding the option for nonlinear trends to the framework approach when appropriate, we gleaned additional insights for Topeka Shiner.

More specific directions for future data collection-analysis and restoration options were gained by including more targeted biotic metrics (Objective 2C). Trends in *total species richness* provided useful and consistent ecological context in our models. However, *species richness*, as an aggregate community metric, is agnostic to species identity and may provide limited directions for research and management intervention relevant to specific effects such as predator stocking, competitive exclusion, or nest habitats (Fletcher Jr. et al., 2025). In our analyses, the sites with Topeka Shiner present or absent substantially overlapped in *total species richness*, suggesting that the identity of fish species in the community may be more relevant to Topeka Shiner presence than the total number of species. Upon subdividing *total species richness* into functional groups, *predator species richness* provided the most informative biotic relationship with Topeka Shiner presence. However, prior Topeka Shiner research has emphasized individual predator species (e.g. Largemouth Bass (*Micropterus nigricans*); Knight & Gido, 2005), rather than *predator species richness* (e.g., the aggregate of Channel Catfish (*Ictalurus punctatus*), Creek Chub (*Semotilus atromaculatus*), Green Sunfish (*Lepomis cyanellus*), Largemouth Bass and other predator species together). Here, choosing to focus our attention on a subset of species (predators) rather than all species or just one predator taxon leads us to stronger and more specific insights for understanding Topeka Shiner presence in Kansas stream communities.

By investigating potential interactions among variables (Objective 2D), more complex drivers of distribution can be identified that can guide sampling, future research, and management actions. Hydrological variables, like *baseflow index*, have been known to interact with other habitat variables (e.g. stream structure, temperature; Rosenfeld, 2017). In our

analyses, we examined and interpreted the effects of *clay* and *baseflow index* on Topeka Shiner persistence together, a novel relationship which may require further assessment in field and theoretical research.

By including random effects (Objective 2E), data analysts can assess generality of whole-state trends help identify relationships that are artifacts of differences among systems rather than clear mechanisms between the variable and species presence. Our systematic approach in previous model iterations simplified analyses but may also have introduced spatial biases across major river basins. Researchers in ecology use random effects to absorb background heterogeneity among groups (here, river basins), increase generality of trends, and reduce the likelihood of spurious correlations (Bolker et al., 2009). Our mixed effects model results suggested that *stream depth*, *clay*, *baseflow index*, and *predator species richness* are robust indicators of Topeka Shiner presence across all Kansas river basins. However, the effect of *number of dams* in earlier models Objectives 1, 2B-D) may not apply across the entire state of Kansas, based on our current dataset. Specifically, the statistical effect of *number of dams* in fixed effect models appears to have been driven by underlying variation across river basins.

Furthermore, the prediction maps from both the mixed effect model (Objective 2E) and the original model (Objective 1) can be used to compare where fish are likely to occur given relationships between previous presences and habitat variables. For clarity, we did not use predictions of fish probability as conjectures that populations should be found at given sites or as absolute guarantees for detection of rare fish at a site if present. Rather, we used the predictions to assess status, probe ecological relationships, and recommend allocation of resources and conservation tools at the request of management partners. This distinction is important for

separating a basic understanding of biogeographic distribution from a practical guide for restoration.

Because statistical analyses of field monitoring data can provide potentially erroneous and inconsistent results, checks to verify species distribution models and variable selection analyses for reliable statewide management applications are especially important for imperiled species. The five connected iterations developed here build upon the original Rode et al. (2025) framework and provide additional ways to check the ecological realism of statistical results. Thus, we gained additional confidence in our analysis of messy field data when we detected similar trends as we systematically repeated model processes with incremental additions.

Take-Home 3: Benefits For Research and Management– Objective 3, Framework Steps 10-12.

By applying the framework and compiling outcomes of Steps 1-9 across five iterations, we identified useful and testable questions regarding the abiotic and biotic variable impact on Topeka Shiner presence (Step 10). To target emerging questions and gaps, in the text that follows, we develop a variety of sampling plans (Step 11). Information generated from our analyses can inform collaborative agency review processes for fish protections (EPA, 2007; USFWS & NMFS, 1986), as well as suggest subsequent framework cycles (Step 12). Below we develop six ideas that integrate knowledge across all iterations of the 12-step framework to generate ecological insights, inform specific conservation actions, and provide recommendations for future data collection and analyses for Topeka Shiner (Objective 3, Framework Steps 10-12).

First, refining how larger scale soil measurements (e.g., *clay*) link to instream water retention is a promising avenue to identify sites where Topeka Shiner might be present or where

they could persist if translocated. *Clay* composition was a consistently positive predictor of Topeka Shiner presence as a function of the surrounding landscape. In the gSSURGO database from which catchment-level soil information was obtained, the *clay* measurements represent surface horizon soil samples on land. Therefore, the coarse metric may or may not reflect conditions in the wetted substrate available to fish; although, *clay* substrates in stream habitat should not preclude the presence of other soil textures more commonly associated with Topeka Shiner life history traits (e.g., gravel). We do not yet fully understand what these landscape scale soil substrate measures mean to Topeka Shiner, but four testable hypotheses may explain which stream sites will support Topeka Shiner. First, prairie stream systems with clay-beds or clay-dominated landscapes may retain more water among summer pool refugia than their sand-bed counterparts as fine materials deposit during low flows (Matthews, 1988). Second, the low permeability of *clay* could be associated with minimizing seepage in the stream bed during low flow conditions in stream systems (Winter et al., 1998). Third, *clay* in the watershed could increase runoff into the stream channel during precipitation events (Waterman et al., 2022). Four, near stream soil geology could contribute to regulating baseflow (e.g. slower infiltration; Price, 2011). Thus, for small stream fish in the Kansas Great Plains, *clay* landscapes could be indicative of streams with geological characteristics that mitigate vertical water loss in critical Topeka Shiner habitat during seasonal and climatic droughts. This raises a key question: “Does *clay* composition in the streambed, riparian connections, or adjacent drainage area affect the persistence of water in stream channels?” If this relationship is biologically meaningful, it could help managers identify promising sites for reintroduction. Subsequent results could support management actions by prioritizing stream segments and systems with geologic characteristics

tied to habitat persistence for developing potential cooperative recovery agreements (KDWP & USFWS, 2023).

Second, linking measures of *baseflow index* to instream, seasonal water could also help researchers and managers prioritize streams with a persistent flow that will support Topeka Shiner presence. *Baseflow index* was a consistent predictor of Topeka Shiner presence, suggesting that this species most often occurs in areas with intermediate levels of hydrologic support. In StreamCat, *baseflow index* is a catchment-scale estimate for the percentage of base flow to stream flow, which assumes that groundwater is a major contributor. However, the *baseflow index* combines multiple sources of sustained flow (e.g. groundwater, soil infiltration, outflow from wetlands or other surface water bodies; Price, 2011) and functions as an indirect proxy for sustained water for fish habitat on average. This information raises the question “Can moderate values of *baseflow index* maintain flow in small prairie systems that have risks of drying?” Operationally, this question could be addressed by measuring real-time wetted conditions in streams using inexpensive water level recorders and relating the flow conditions that fish experience to *baseflow index* across a gradient of conditions (soil moisture contents, water-table elevation, groundwater spring presence). These data could support management actions to protect critical Topeka Shiner habitat from water extraction and dewatering through water-allocation decisions (Kansas Legislature, 1984).

Third, the interaction between *clay* and *baseflow index* further suggests that Topeka Shiner may respond to multiple combinations of geologic and hydrologic conditions. Each could reflect a habitat need for Topeka Shiner of stable refugia in small intermittent systems that are prone to drying. Both broad-scale metrics may be relevant to drought refugia and water persistence in small prairie streams, in that *clay* influences water loss and *baseflow index*

corresponds to water inputs. This raises the question of “Can *clay*-dominated soils in the surface geology compensate for low base flow?” To address this question, data can be collected on how different combinations of *clay* and *baseflow index* affect Topeka Shiner presence or absence.

Fourth, combining easily measured main channel *stream depth* metrics with other measures of stream size could create multi-metric indicators that identify small systems that can support Topeka Shiner. In our analyses, *stream depth* was a consistently negative predictor of Topeka Shiner presence as a function of the main channel habitat. Our metric for *stream depth* was estimated as a function of the entire drainage area (NHDPlus, Bieger et al., 2015), and is a measure of size of the entire sample location rather than specific habitat conditions encountered by fish at the time of sampling. As an advantage, we were able to acquire *stream depth* data at all fish sampling sites. Agency sampling protocols typically include site-specific measurements of water depth along one or more transects, which more directly represent available habitat for fish in the channel and at the riparian edge (e.g. KDWP, 2008), but these measurements are too labor intensive to undertake at all fish sampling sites statewide. These results raise the question of “Can multiple morphological metrics be combined to characterize the system size that supports Topeka Shiner persistence across a wide range of systems?” This question could be addressed by linking the broad GIS *stream depth* patterns estimated in our global model with local-scale measurements of fish-specific, site-specific, and mesohabitat-specific depths in the thalweg, within available mesohabitat (pools, runs, etc.), then cross-referencing locations where Topeka Shiner is present or absent. Such a link could support remediation recommendations associated with permit review and could help prioritize habitat restoration efforts in altered channels (KDWP & USFWS, 2023).

Fifth, while prevailing water availability issues in Kansas are critical (Dodds et al., 2004), we also identify future directions for other threats (e.g. local predation and artificial barriers; Gerken & Paukert, 2013; Pfaff et al., 2023). In our analysis, *predator species richness* was a consistently negative predictor of Topeka Shiner presence as a function of riverscape community composition. In our dataset, *predator species* richness was based on species presence-absence, so it did not capture relative body-sizes, abundances, or life stages, all of which influence whether observed predator populations can adversely affect Topeka Shiner. This leads us to ask the question of “Can more detailed assessments of predators, competitors, and nest builders (e.g., abundance, age-size classes, seasonal behaviors) help identify specific biotic pressures that influence Topeka Shiner presence or absence?” These guild-based analyses can guide recommendations for sportfish predator stocking plans by identifying priority release locations or timing that minimize predation risk (Campbell et al., 2016; USFWS, 2024).

Sixth, we did not find consistent evidence that the *number of dams* in mainstem streams predicts Topeka Shiner occurrence, suggesting that the effect of such barriers remains uncertain. The metric used here reflects dam counts of mostly large dams with specific heights and storage capacities (> 7.62 m with 1.85 ha-m, or 1.83 m with 6.17 ha-m; Cooper & Infante, 2024), at locations that may be less relevant to the upstream reaches where Topeka Shiner occur (e.g. Schrank et al., 2001). This raises the question “Do dam metrics indicate coarse levels of stream fragmentation, regional differences, or direct barrier effects?” For management agencies, this means that instream barriers may still matter, but the available data with gaps in intermediate dam counts or with large dams alone cannot adequately detect trends at the statewide level. If clarified, understanding the role of dams in Topeka Shiner distributions in Kansas could support

management decisions for dam removal, mitigation, or construction projects (Kansas Legislature, 1917; USFWS, 2024).

Conclusions - Advantages of Our Process.

Using this framework (original and adapted iterations) can improve the utility of existing monitoring data for Topeka Shiner by providing a step-by-step, quantitative method to process messy data that often violates many statistical assumptions. The framework is different than the one-time analyses that most often occur for Topeka Shiner research and monitoring. Our study illustrates how intentionally connecting multiple cycles of the framework provides checkpoints to evaluate uncertainties and inconsistencies that may arise. The consistency of our results for *system size*, *substrate*, *flow*, and *biota* demonstrates how our application of the framework unearths new directions for Topeka Shiner research, management, and conservation and demonstrates the potential to support similar analyses of monitoring data for other rare species.

Acknowledgements

The project was administered by the Kansas Cooperative Fish and Wildlife Research Unit as a joint effort between Kansas State University, the U.S. Geological Survey, U.S. Fish and Wildlife Service, the Kansas Department of Wildlife and Parks (KDWP), and the Wildlife Management Institute. We thank KDWP for sharing the monitoring dataset. We thank Michael Madin and Jean Francois from the Kansas State Department of Geography and Geospatial Analysis who have offered their expertise in R programming and spatial data analysis. Jordan Hoffmeier provided valuable insights into conservation concerns highlighted within the Kansas SWAP as well as substantial expertise in aquatic ecology. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. A review of scientific and common names was completed. Tribe and geologic names do not apply to this manuscript.

Tables

Table 2.1. Variables within 8 ecological categories used to model Topeka Shiner distributions in Kansas. Included are variable name, relevance to fish, definition of measurement, units, spatial scale, database used, and source.

Ecological Variable	Metric	Fish Relevance	Definition	Units	Spatial Scale	Database	Source
Stream Size	<i>Stream Depth</i>	Stream depth influences flow regimes and structures fish communities. Shallow streams are associated with less flow and smaller communities, while deeper areas have larger quantities of flow and fish species.	Estimated depth for flowline exit point.	Meters	NHDPlus Flowlines	National Hydrography Data Plus V2 (NHDPlus)	McKay et al., 2012
Stream Size	<i>Stream Order</i>	Stream order is a common description of a stream's position within the river network, with values increasing downstream of confluences. Stream size or order can influence fish distributions, with some fish occupying a wide range (1-9) and others restricted to a few orders.	Modified Strahler Stream Order	Number	NHDPlus Flowlines	NHDPlus	McKay et al., 2012
Substrate	<i>Clay</i>	Soil composition of the stream channel and adjacent catchment areas may influence bed substrate and regulate runoff and baseflow delivery to streams.	Calculated average Soil Surface Horizon A Composition, ratio of mass of clay (<0.002mm) to mass of soil fines.	Percent	NHDPlus Catchments	Soil Survey Geographic Database (SSURGO)	NRCS, 2022
Flow	<i>Stream Flow</i>	Stream flow is often measured as water velocity, or the rate of water movement through a system. Human alterations to the flow regime can be detrimental to many aspects of fish life history.	Gage-adjusted mean annual (1971-2000) velocity for flowline exit point (Jobson method)	Feet Per Second	NHDPlus Flowlines	NHDPlus	McKay et al., 2012
Flow	<i>Baseflow Index</i>	Base flow is the water flowing in stream channel excluding runoff, partially attributed to groundwater. Human water extraction can reduce water available for fish with minimum flow needs.	Calculated index for the contribution of groundwater and other sustained sources to total stream flow	Percent	NHDPlus Catchments	The Stream Catchment Database (StreamCat)	Hill et al., 2016

Table 2.1. Variables within 8 ecological categories used to model Topeka Shiner distributions in Kansas. Included are variable name, relevance to fish, definition of measurement, units, spatial scale, database used, and source.

Ecological variable	Metric	Fish Relevance	Definition	Units	Spatial Scale	Database	Source
Biota	<i>Species Richness</i>	Species richness is commonly used as a measure of ecosystem health and biotic community size.	Calculated from all fish species present in one stream sample	Number	Site Sample	Kansas Department of Wildlife & Parks (KDWP)	Not Published
Land Use	<i>Grassland</i>	Grassland cover is associated with prairie and plains ecosystems. Kansas' historical grasslands are associated with co-evolved stream fish communities.	Calculated grassland cover in 2021	Percent	NHDPlus Catchments	National Land Cover Database (NLCD)	Dewitz, 2023
Gradient	<i>Gradient</i>	Channel slope, or gradient, is a commonly measured fish habitat variable. A reaches gradient can dictate how quickly the river flows, and therefore what habitat types and fauna are present.	Slope of flowline, calculated as the change in elevation over the length of the flowline.	Meters/Meters	NHDPlus Flowlines	NHDPlus	McKay et al., 2012
Sinuosity	<i>Curvature</i>	Curvature is a measure of stream channel sinuosity and is associated with microhabitat diversity. Channelization, flow alteration, and bank stabilization for human use reduces sinuosity and straightens river paths.	Calculated average geometric curvature from flowline points at least 60m apart.	1/Meters	NHDPlus Flowlines	NHDPlus	McKay et al., 2012
Land Use	<i>Development</i>	Human development such as urbanization can lead to changes in water quality and channel structure of streams. Therefore, developed land use of varying intensities could alter fish distribution.	Calculated developed land use cover (Open space, low-high intensities) for 2021	Percent	NHDPlus Catchments	NLCD	Dewitz, 2023

Table 2.1. Variables within 8 ecological categories used to model Topeka Shiner distributions in Kansas. Included are variable name, relevance to fish, definition of measurement, units, spatial scale, database used, and source.

Ecological variable	Metric	Fish Relevance	Definition	Units	Spatial Scale	Database	Source
Land Use	<i>Cropland</i>	Agricultural practices, including land used for crop production, can potentially alter nearby stream systems through water extraction or sediment runoff. Stream habitat alterations (flow, turbidity, etc.) can lead to declines in fish populations.	Calculated agricultural cropping land use cover in 2021	Percent	NHDPlus Catchments	NLCD	Dewitz, 2023
Land Use	<i>Pastureland</i>	Agricultural practices, including land used for grazing or haying, can potentially alter nearby stream systems through nutrient runoff. Stream habitat alterations (productivity, algal blooms, etc.) can lead to declines in fish populations.	Calculated agricultural pastureland use cover in 2021	Percent	NHDPlus Catchments	NLCD	Dewitz, 2023
Dams	<i>Number of Dams</i>	Human construction of dams can alter the stream habitat, water quality, and system connectivity. Dams can change fish assemblages by providing more lentic habitat for some fish and removing lotic habitat for others.	Total mainstem dam count (>7.62m)	Number	NHDPlus Flowlines	National Anthropogenic Barriers Dataset (NABD) & NHDPlus	Cooper & Infante, 2024; Ostroff et al., 2013

Table 2.2. Conceptual progression of framework iterations examining Topeka Shiner distributions in Kansas by research objective, model iteration (Steps 7-9), future directions (Steps 10-11), and applications (Steps 11-12). Included are summaries of research questions, motivations, predictions, results, emerging questions, proposed future measurements, priority locations, research importance, and management implications associated with each objective. New questions addressed by existing data have “not applicable” (N/A) listed in remaining sections. Examples of future directions and applications provided are non-exhaustive.

Objective	Model Iteration	Future Direction	Application
1	Question: What environmental variables are linked to the site-specific probability of Topeka Shiner presence or absence across Kansas? Motivation: Topeka Shiner is a high-priority threatened species identified in the State Wildlife Action Plan. Prediction: Environmental variables identified in the range-wide literature review would be associated with Topeka Shiner distributions. Results: Increasing <i>clay</i> content and decreasing <i>stream depth</i>, <i>species richness</i>, and <i>number of dams</i> were significantly associated with Topeka Shiner presence.	New Questions: • How do data visualizations and modeled trends agree or disagree regarding environmental influences on Topeka Shiner distributions? • What groups of species explain the negative relationship with <i>species richness</i>? • Does <i>clay</i> composition, riparian connectivity, or drainage area affect stream water persistence? • Can multiple morphological metrics characterize system size suitable for Topeka Shiner persistence? Future Measurements: <i>Clay</i> composition of surface soils, bed substrate, and bedrock; water residence time; pool and thalweg depth across stream orders.	Priority Locations: Areas with high predicted occurrence and >50% <i>clay</i> soil composition, particularly Mill Creek tributaries. Research Importance: Ground-truth the <i>clay</i>-soil relationship as a mechanism mediating drought resistance; reconcile scale-dependent patterns. Management Implications: Identify habitat restoration opportunities and potential reintroduction sites.
2A	Question: How do data visualizations and modeled trends agree or disagree regarding environmental influences on Topeka Shiner distributions? Motivation: Multiple lines of quantitative evidence increase confidence in model interpretation. Prediction: Similarities and differences would emerge across analytical approaches, with agreement on well-established habitat variables and disagreement on weaker or nonlinear relationships.	New Questions: • Does <i>baseflow index</i> influence Topeka Shiner distributions as a nonlinear variable? Future Measurements: N/A	Priority Locations: N/A Research Importance: Evaluate consistency among analytical approaches. Management Implications: N/A

Table 2.2. Conceptual progression of framework iterations examining Topeka Shiner distributions in Kansas by research objective, model iteration (Steps 7-9), future directions (Steps 10-11), and applications (Steps 11-12). Included are summaries of research questions, motivations, predictions, results, emerging questions, proposed future measurements, priority locations, research importance, and management implications associated with each objective. New questions addressed by existing data have “not applicable” (N/A) listed in remaining sections. Examples of future directions and applications provided are non-exhaustive.

Objective	Model Iteration	Future Direction	Application
2A	Results: Most environmental variables showed agreement between plots and models; <i>baseflow index</i> was a notable exception.		
2B	Question: Does <i>baseflow index</i> influence Topeka Shiner distributions as a nonlinear variable? Motivation: Conflicting patterns for <i>baseflow index</i> were identified in Objective 2A. Prediction: Topeka Shiner occurrence would peak at intermediate <i>baseflow index</i> values. Results: A nonlinear (quadratic) relationship best described Topeka Shiner occurrence. <i>Baseflow index</i> became a significant predictor in multiple logistic regression models.	New Questions: • Do environmental variables interact in influencing Topeka Shiner distributions? • Can moderate base flow maintain flow in small prairie streams with drying risk? Future Measurements: Base flow, water velocity, water-table elevation, groundwater spring presence, wetted stream area, and <i>stream depth</i>.	Priority Locations: Areas with high occurrence probability and <i>baseflow index</i> values of approximately 20–30, including Mill Creek and Wildcat Creek. Research Importance: Ground-truth the baseflow relationship as an indicator of habitat permanence. Management Implications: Support water-allocation decisions affecting critical habitat.
2C	Question: What groups of species provide more specific insights about the negative relationship with high <i>species richness</i>? Motivation: Better understanding of functional interactions within fish communities can improve management actions. Prediction: Trait- or interaction-based subsets of the fish community would be more informative than total <i>species richness</i>. Results: <i>Predator species richness</i> most closely mirrored total <i>species richness</i> and was the most explanatory biotic metric.	New Questions: • Can detailed assessments of predator, competitor, and nest-builder abundance, size structure, and seasonal behavior identify specific biotic pressures influencing Topeka Shiner occurrence? Future Measurements: Relative body sizes, abundance, life stages, and seasonal dynamics of Topeka Shiner and co-occurring species.	Priority Locations: Areas with high predicted occurrence and variable <i>species richness</i>, including Deep Creek and Pillsbury Wildlife Crossing. Research Importance: Clarify ecological relationships among co-occurring species. Management Implications: Guide timing and locations for predator stocking; identify suitable reintroduction sites associated with positive species interactions.

Table 2.2. Conceptual progression of framework iterations examining Topeka Shiner distributions in Kansas by research objective, model iteration (Steps 7-9), future directions (Steps 10-11), and applications (Steps 11-12). Included are summaries of research questions, motivations, predictions, results, emerging questions, proposed future measurements, priority locations, research importance, and management implications associated with each objective. New questions addressed by existing data have “not applicable” (N/A) listed in remaining sections. Examples of future directions and applications provided are non-exhaustive.

Objective	Model Iteration	Future Direction	Application
2D	Question: Do multiple environmental variables interact to influence Topeka Shiner distributions simultaneously? Motivation: <i>Baseflow index</i> exhibited a non-constant relationship with occurrence. Prediction: Hydrologic influences would vary depending on physical stream habitat characteristics. Results: <i>Baseflow index</i> and <i>clay</i> content significantly interacted in shaping predicted Topeka Shiner distributions.	New Questions: • Are observed environmental relationships robust despite limited data distribution? • Can <i>clay</i>-rich soils compensate for low baseflow conditions? Future Measurements: Potential manipulative studies incorporating soil and substrate composition, soil moisture, temperature, wetted stream area, <i>stream depth</i>, baseflow, and precipitation.	Priority Locations: Areas with <20% <i>baseflow index</i> and >50% <i>clay</i> soil composition, including Walnut Creek and southern Mill Creek tributaries. Research Importance: Test hydrogeological mechanisms within stream fish ecology. Management Implications: Develop new habitat restoration strategies.
2E	Question: Are trends with environmental variables stable despite the distribution of our limited data? Motivation: Spatial clustering of Topeka Shiner occurrences could bias statewide inference. Prediction: Accounting for regional variability would improve environmental relationship estimates. Results: <i>Baseflow index</i>, <i>clay</i> content, <i>stream depth</i>, and <i>species richness</i> remained significant predictors. Dam effects became unstable after accounting for river-basin random effects. Explanatory and predictive performance improved.	New Questions: • What can be learned by comparing all models developed across Objectives 1–2E? • Do dam metrics reflect habitat fragmentation, regional differences, or direct barrier effects? Future Measurements: Dam size, function, and location; effects of 5–15 dam densities; regional comparisons; movement distances and seasonal migrations of Topeka Shiner.	Priority Locations: Areas containing 5–15 dams that have not been surveyed within the last 20 years. Research Importance: Evaluate mechanisms underlying potential dam-related threats. Management Implications: Inform dam removal, mitigation, or construction decisions.
2F	Question: What do we learn from comparing results across all models developed in Objectives 1–2E? Motivation: Consistencies and discrepancies among models provide an additional weight of evidence beyond individual analyses.	New Questions: N/A Future Measurements: N/A	Priority Locations: N/A Research Importance: Synthesize evidence across modeling approaches and assess robustness of conclusions.

Table 2.2. Conceptual progression of framework iterations examining Topeka Shiner distributions in Kansas by research objective, model iteration (Steps 7-9), future directions (Steps 10-11), and applications (Steps 11-12). Included are summaries of research questions, motivations, predictions, results, emerging questions, proposed future measurements, priority locations, research importance, and management implications associated with each objective. New questions addressed by existing data have “not applicable” (N/A) listed in remaining sections. Examples of future directions and applications provided are non-exhaustive.

Objective	Model Iteration	Future Direction	Application
2F	Prediction: Similarities and differences would emerge in variable importance, trend direction, and model performance metrics. Results: As model complexity increased, the relative influence of <i>clay</i> decreased while the influence of <i>stream depth</i> and <i>baseflow index</i> increased. The most influential variables were generally consistent across models, whereas disagreements primarily reflected modeling structure or weaker relationships.		Management Implications: Increase confidence in prioritizing environmental drivers for conservation planning.

Table 2.3. Objective 1 – Step 8F. Results from multiple logistic regression on standardized data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas ($n=335$) from 670 stream site samples collected between 1947 and 2020. Included are variable name, coefficient (β), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor variables include *stream depth* (estimated in meters), *clay* (% in catchment surface geology), *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *species richness* (number of fish species in a single sample), *grassland* (% cover in catchment), gradient (m/m), *curvature*, (H) *development* (% cover in the catchment), *cropland* (% cover in the catchment), *pastureland* (% cover in the catchment), and *number of dams* (total in the mainstem). These calculations correspond to Step 8F of the Rode et al. 2025 framework.

Variable	β	SE	Z	95% CI		p		OR	VIF
				Lower	Upper				
(Intercept)	-1.35	0.38	-3.52	-2.10	-0.60	0.00	*	0.26	
<i>Stream Depth</i>	-1.86	0.67	-2.80	-3.17	-0.56	0.01	*	0.16	1.77
<i>Clay</i>	2.07	0.26	7.84	1.55	2.59	0.00	*	7.91	1.30
<i>Baseflow Index</i>	-0.24	0.20	-1.18	-0.63	0.16	0.24		0.79	1.30
<i>Species Richness</i>	-0.69	0.14	-4.87	-0.96	-0.41	0.00	*	0.50	1.28
<i>Grassland</i>	0.72	0.41	1.75	-0.09	1.54	0.08		2.06	9.91
<i>Gradient</i>	0.03	0.16	0.17	-0.29	0.35	0.87		1.03	2.16
<i>Curvature</i>	0.12	0.19	0.62	-0.25	0.49	0.54		1.12	1.17
<i>Development</i>	-1.08	0.66	-1.64	-2.37	0.21	0.10		0.34	1.38
<i>Cropland</i>	0.09	0.29	0.31	-0.48	0.66	0.75		1.09	6.31
<i>Pastureland</i>	-0.24	0.22	-1.13	-0.67	0.18	0.26		0.78	2.84
<i>Number of Dams</i>	-0.79	0.16	-4.84	-1.11	-0.47	0.00	*	0.45	1.78

Table 2.4. *Objective 2B – Step 8F.* Results from multiple logistic regression on standardized data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas ($n=335$) from 670 stream site samples collected between 1947 and 2020, with an additional parameter added to describe a quadratic relationship with *baseflow index* (underlined). Included are variable name, coefficient (β), standard error (SE), Z-statistic (Z), lower and upper 95% confidence interval, p -value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor variables include *stream depth* (estimated in meters), *clay* (% in catchment surface geology), *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *species richness* (number of fish species in a single sample), *grassland* (% cover in catchment), *gradient* (m/m), *curvature*, *development* (% cover in the catchment), *cropland* (% cover in the catchment), *pastureland* (% cover in the catchment), and *number of dams* (total in the mainstem). All other methodological details and variables are the same as described in Objective 1. The results in this table correspond to Fig. 8. These calculations correspond to Step 8F of the Rode et al 2025 framework.

Variable	β	SE	Z	95% CI		p		OR	VIF
				Lower	Upper				
(Intercept)	-0.66	0.40	-1.67	-1.44	0.12	0.10		0.52	
<i>Stream Depth</i>	-1.78	0.65	-2.75	-3.04	-0.51	0.01	*	0.17	1.75
<i>Clay</i>	1.44	0.29	4.99	0.87	2.00	0.00	*	4.22	1.39
<i>Baseflow Index</i>	-0.57	0.42	-1.36	-1.41	0.26	0.17		0.56	2.11
<u><i>Baseflow Index</i>²</u>	<u>-1.56</u>	<u>0.47</u>	<u>-3.29</u>	<u>-2.49</u>	<u>-0.63</u>	<u>0.00</u>	<u>*</u>	<u>0.21</u>	<u>2.35</u>
<i>Species Richness</i>	-0.67	0.15	-4.56	-0.95	-0.38	0.00	*	0.51	1.28
<i>Grassland</i>	0.70	0.41	1.72	-0.10	1.51	0.09		2.02	8.97
<i>Gradient</i>	-0.11	0.17	-0.66	-0.44	0.22	0.51		0.90	2.16
<i>Curvature</i>	0.11	0.20	0.53	-0.28	0.50	0.59		1.11	1.22
<i>Development</i>	-1.27	0.70	-1.82	-2.64	0.10	0.07		0.28	1.39
<i>Cropland</i>	0.06	0.29	0.20	-0.52	0.63	0.85		1.06	5.91
<i>Pastureland</i>	-0.05	0.22	-0.22	-0.47	0.38	0.83		0.95	2.76
<i>Number of Dams</i>	-0.90	0.18	-5.06	-1.24	-0.55	0.00	*	0.41	1.97

Table 2.5. *Objective 2C – Step 8F.* Results from multiple logistic regression on standardized data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas ($n=335$) from 670 stream site samples collected between 1947 and 2020, with predator species richness (underlined). Included are variable name, coefficient (β), standard error (SE), Z-statistic (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor variables include *stream depth* (estimated in meters), *clay* (% in catchment surface geology), *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *predator species richness* (number of fish species in a single sample which are potential predators of Topeka Shiner), *grassland* (% cover in catchment), *gradient* (m/m), *curvature*, *development* (% cover in the catchment), *cropland* (% cover in the catchment), *pastureland* (% cover in the catchment), and *number of dams* (total in the mainstem). All other methodological details and variables are the same as described in Objective 1. The results in this table correspond to Fig. 10. These calculations correspond to Step 8F of the Rode et al. 2025 framework.

Variable	β	SE	Z	95% CI		p		OR	VIF
				Lower	Upper				
(Intercept)	-0.72	0.41	-1.76	-1.53	0.08	0.08		0.48	
<i>Stream Depth</i>	-1.92	0.68	-2.82	-3.25	-0.59	0.00	*	0.15	1.78
<i>Clay</i>	1.41	0.29	4.83	0.84	1.98	0.00	*	4.09	1.38
<i>Baseflow Index</i>	-0.64	0.43	-1.50	-1.47	0.20	0.13		0.53	2.13
<i>Baseflow Index²</i>	-1.54	0.47	-3.26	-2.46	-0.62	0.00	*	0.21	2.35
<u><i>Predator Species Richness</i></u>	<u>-0.73</u>	<u>0.14</u>	<u>-5.08</u>	<u>-1.01</u>	<u>-0.45</u>	<u>0.00</u>	<u>*</u>	<u>0.48</u>	<u>1.29</u>
<i>Grassland</i>	0.85	0.43	1.99	0.01	1.68	0.05	*	2.33	9.39
<i>Gradient</i>	-0.11	0.17	-0.68	-0.44	0.21	0.50		0.89	2.17
<i>Curvature</i>	0.12	0.20	0.58	-0.28	0.51	0.56		1.12	1.23
<i>Development</i>	-1.21	0.72	-1.68	-2.62	0.21	0.09		0.30	1.41
<i>Cropland</i>	0.15	0.30	0.48	-0.45	0.74	0.63		1.16	6.24
<i>Pastureland</i>	-0.02	0.22	-0.10	-0.46	0.41	0.92		0.98	2.82
<i>Number of Dams</i>	-0.98	0.18	-5.45	-1.34	-0.63	0.00	*	0.37	2.02

Table 2.6. Objective 2D – Step 8F. Results from multiple logistic regression on standardized data using 5 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas ($n=335$) from 670 stream site samples collected between 1947 and 2020, with an interaction term included between *baseflow index* and *clay*. Eighty one percent of the data was collected between 2000-2021. Included are variable name, coefficient (β), standard error (SE), Z-statistic (Z), lower and upper 95% confidence interval, p -value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor variables include *stream depth* (estimated in meters), *clay* (% in catchment surface geology), *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *predator species richness* (number of fish species in a single sample which are potential predators of Topeka Shiner), and *number of dams* (total in the mainstem). All two and three way interactions were tested but only the interaction shown below is significant. The reduced model excludes noncontributing variables. All other methodological details and variables are the same as described in Objective 1. The new variable metric added is underlined. The results in this table correspond to Fig. 11. These calculations correspond to Step 8F of the Rode et al. 2025 framework.

Variable	β	SE	Z	95% CI		p		OR	VIF
				Lower	Upper				
(Intercept)	-0.44	0.38	-1.16	-1.18	0.30	0.24		0.64	
<i>Stream Depth</i>	-2.60	0.64	-4.09	-3.84	-1.35	0.00	*	0.07	1.71
<i>Clay</i>	1.35	0.28	4.91	0.81	1.89	0.00	*	3.87	1.27
<i>Baseflow Index</i>	-0.28	0.35	-0.80	-0.97	0.41	0.42		0.75	1.29
<i>Baseflow Index</i> ²	-1.76	0.45	-3.87	-2.65	-0.87	0.00	*	0.17	1.48
<i>Predator Species Richness</i>	-0.65	0.14	-4.74	-0.92	-0.38	0.00	*	0.52	1.56
<i>Number of Dams</i>	-0.64	0.15	-4.37	-0.93	-0.35	0.00	*	0.53	2.35
<u><i>Clay: Baseflow Index</i></u>	<u>-1.09</u>	<u>0.56</u>	<u>-1.96</u>	<u>-2.18</u>	<u>0.00</u>	<u>0.05</u>	<u>*</u>	<u>0.34</u>	<u>1.37</u>

Table 2.7. Objective 2E – Step 8F. Results from multiple logistic regression on standardized data using 5 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas ($n=670$), with a random effect of major river basins. Eighty one percent of the data was collected between 2000-2021. Included are variable name, coefficient (β), standard error (SE), Z-statistic (Z), lower and upper 95% confidence interval, p -value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF) for fixed effects and standard deviation (SD) and lower and upper 95% confidence interval for random intercepts. Predictor variables include *stream depth* (estimated in meters), *clay* (% in catchment surface geology), *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *predator species richness* (number of fish species in a single sample which are potential predators of Topeka Shiner), and *number of dams* (total in the mainstem). All other methodological details and variables are the same as described in Objective 2D. The results in this table correspond to Fig. 13. These calculations correspond to Step 8F of the Rode et al. 2025 framework.

		95% CI								
	Variable	β	SE	Z	Lower	Upper	p	OR	VIF	
Fixed Effects	(Intercept)	-7.84	3.96	-1.98	-15.60	-0.08	0.05	0.00		
	<i>Stream Depth</i>	-2.75	0.60	-4.57	-3.93	-1.57	0.00	*	0.06	1.30
	<i>Clay</i>	0.85	0.33	2.61	0.21	1.50	0.01	*	2.35	1.45
	<i>Predator Species Richness</i>	-0.68	0.15	-4.50	-0.98	-0.39	0.00	*	0.51	1.05
	<i>Baseflow Index</i>	-0.07	0.44	0.16	-0.80	0.93	0.88		1.07	1.41
	<i>Baseflow Index</i> ²	-1.57	0.56	-2.82	-2.67	-0.48	0.00	*	0.21	2.09
	<i>Number of Dams</i>	0.77	1.34	0.58	-1.85	3.40	0.56		2.16	1.15
	<i>Clay: Baseflow Index</i>	-1.78	0.74	-2.42	-3.22	-0.34	0.02	*	0.17	1.93
		95% CI								
		SD	Lower		Upper					
Random Intercept	Basin	7.21	1.8		28.83					

Table 2.8. Objective 2F – Step 8F. Comparison of standardized estimate magnitudes from candidate multiple logistic regression models predicting the probability of Topeka Shiner presence in Kansas ($n= 335$) from 670 stream site samples collected between 1947 and 2020. Included are top models for Objectives 1-2E, corrected Akaike Information Criterion (AIC_c), variable names, beta coefficients (β), statistical significance (asterisk to denote $p \leq 0.05$), and relative ranks of significant variables within models (1 = highest magnitude of influence on predicted Topeka Shiner presence). Variables shown include *stream depth* (estimated in meters), *clay* (% in catchment surface geology), *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), *species richness* (number of fish species in a single sample), *predator species richness* (number of fish species in a single sample which are potential predators to Topeka Shiner), *grassland* (% cover in catchment), *gradient* (m/m), *curvature*, *development* (% cover in the catchment), *cropland* (% cover in the catchment), *pastureland* (% cover in the catchment), and *number of dams* (total in the mainstem). Blanks (--) for terms not included in models. All β are relative to the individual model intercept.

ΔAIC_c	Best Model									
	Obj 1		Obj 2B		Obj 2C		Obj 2D		Obj 2E	
	50.28		27.57		22.49		41.59		0	
β /Rank										
<i>Stream Depth</i>	-1.86*	2	-1.78*	2	-1.92*	2	-2.60*	2	-2.75*	2
<i>Clay</i>	2.07*	1	1.44*	3	1.41*	3	1.35*	3	0.85*	4
<i>Baseflow Index</i>	-0.24		-0.57		-0.64		-0.28		-0.07	
<i>Baseflow Index²</i>	--		-1.56*	1	-1.54*	1	-1.76*	1	-1.57*	1
<i>Species Richness</i>	-0.69*	4	-0.67*	5	--		--		--	
<i>Predator Species Richness</i>	--		--		-0.73*	6	-0.65*	5	-0.68*	
<i>Grassland</i>	0.72		0.70		0.85*	5	--		--	
<i>Gradient</i>	0.03		-0.11		-0.11		--		--	
<i>Curvature</i>	0.12		0.11		0.12		--		--	
<i>Development</i>	-1.08		-1.27		-1.21		--		--	
<i>Cropland</i>	0.09		0.06		0.15		--		--	
<i>Pastureland</i>	-0.24		-0.05		-0.02		--		--	
<i>Number of Dams</i>	-0.79*	3	-0.90*	4	-0.98*	4	-0.64*	6	0.77	
<i>Clay: Baseflow Index</i>	--		--		--		-1.09*	4	-1.78*	3

Figures

Figure 2.1. Modified Rode et al. (2025) adaptive management framework showing the objectives addressed here. In our first objective, we apply the original framework to the Topeka Shiner (*Notropis topeka*), using a statewide Kansas monitoring database. The 12 steps are described in the text. In the second objective, we test six data-driven questions in multiple iterations of the 12-step cycle. These connected iterations include (A) comparing data to identify agreement and disagreement in trends, (B) testing nonlinear relationships, (C) refining ecological variable metrics, (D) testing interactions between ecological variables, (E) incorporating random effects, and (F) comparing statistical models to identify consistently influential variables. For our third objective, we integrate each iteration of the 12-step framework to develop ecological insights and recommendations for future data collection and conservation actions (discussion).

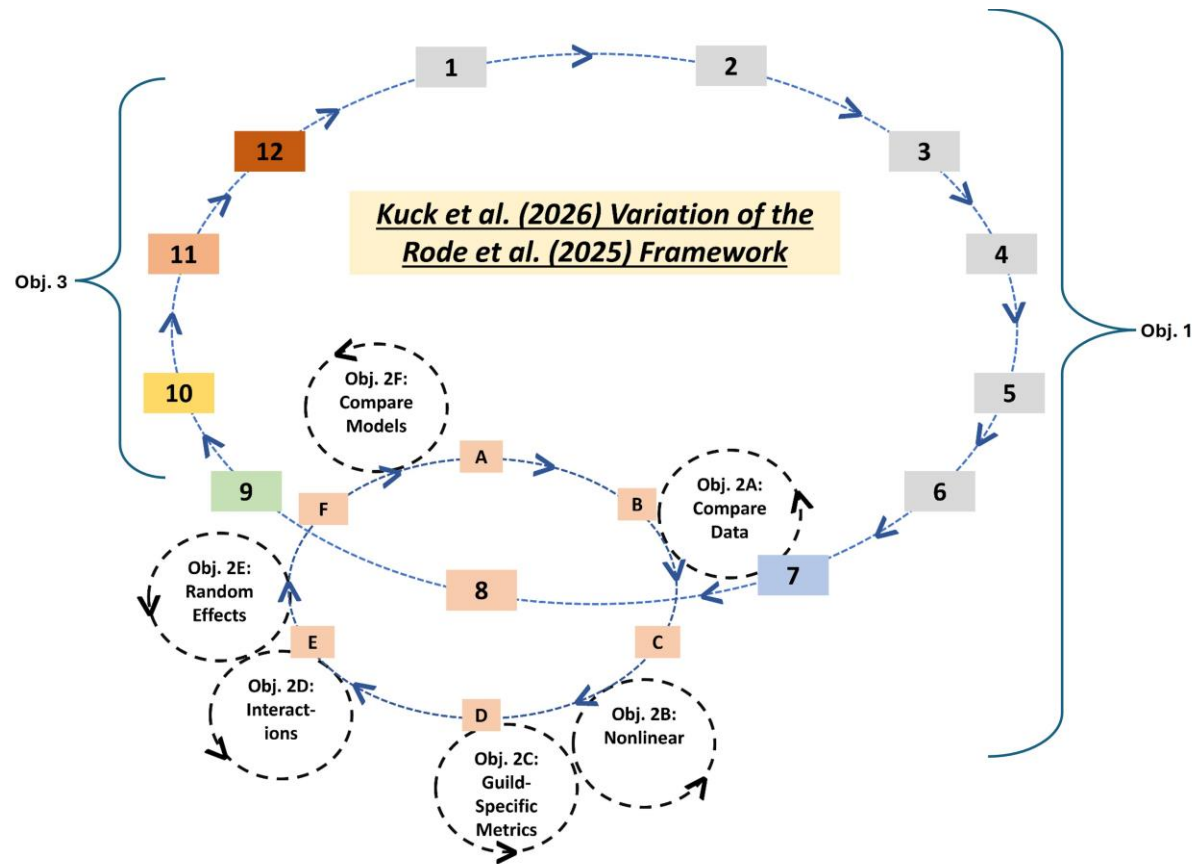


Figure 2.2. *Objective 1 – Step 3.* Number of instances in Topeka Shiner literature review of 35 articles where the ecological categories of (A) system size, (B) substrate, (C) flow, (D) biota, (E) land use, and (F) dams were discussed or investigated in Kansas and other states. Percent of statistically significant ($p \leq 0.05$; +/-) and nonsignificant (NS; including qualitative or untested) trends included below with letters (*a- ah) corresponding to citations in Appendix Table 2.

WOS Literature Review ($N = 35$ articles)

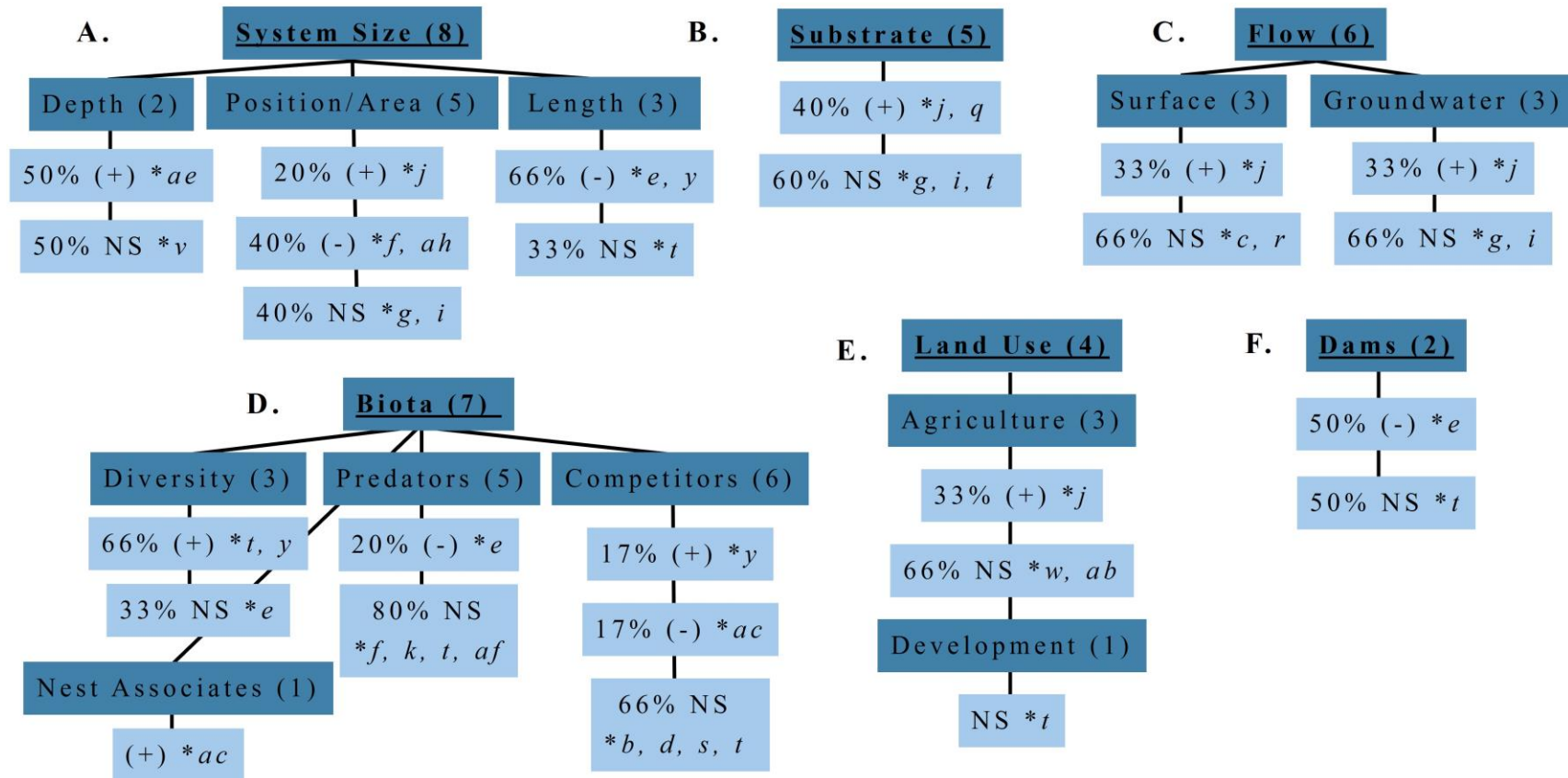


Figure 2.3. *Objective 1 – Step 8A.* Map of the major river basins in Kansas (numbered 1-12) with points to represent stream biodiversity sample sites ($N=12,086$) collected between 1857 and 2021. Eighty-one percent of data were collected between 2000-2021. Topeka Shiner was present at sites labeled with red points ($n=341$), while the fish was absent among sites in blue.

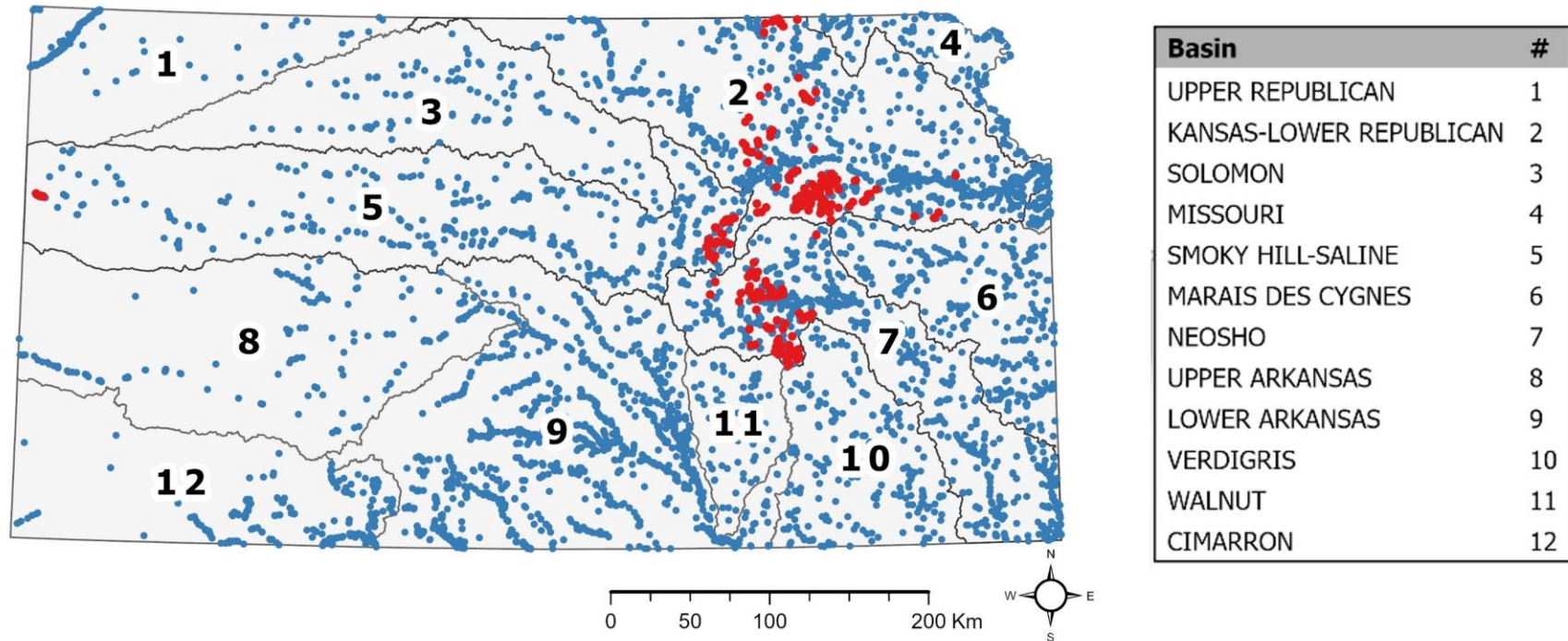


Figure 2.4. Objective 1 – Step 8C. Kernel Density Estimate (KDE) ridge plots of eleven ecological variables (X axis) calculated from the original data. Variables shown include (A) *stream depth* (estimated in meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). The Y Axis is the probability density of a Topeka Shiner presence ($n = 335$; red) or absence (blue) sample occurring at a given value of ecological variables. Data were collected from 670 stream sites in Kansas from 1947 to 2020. The area under each curve is equal to 1.

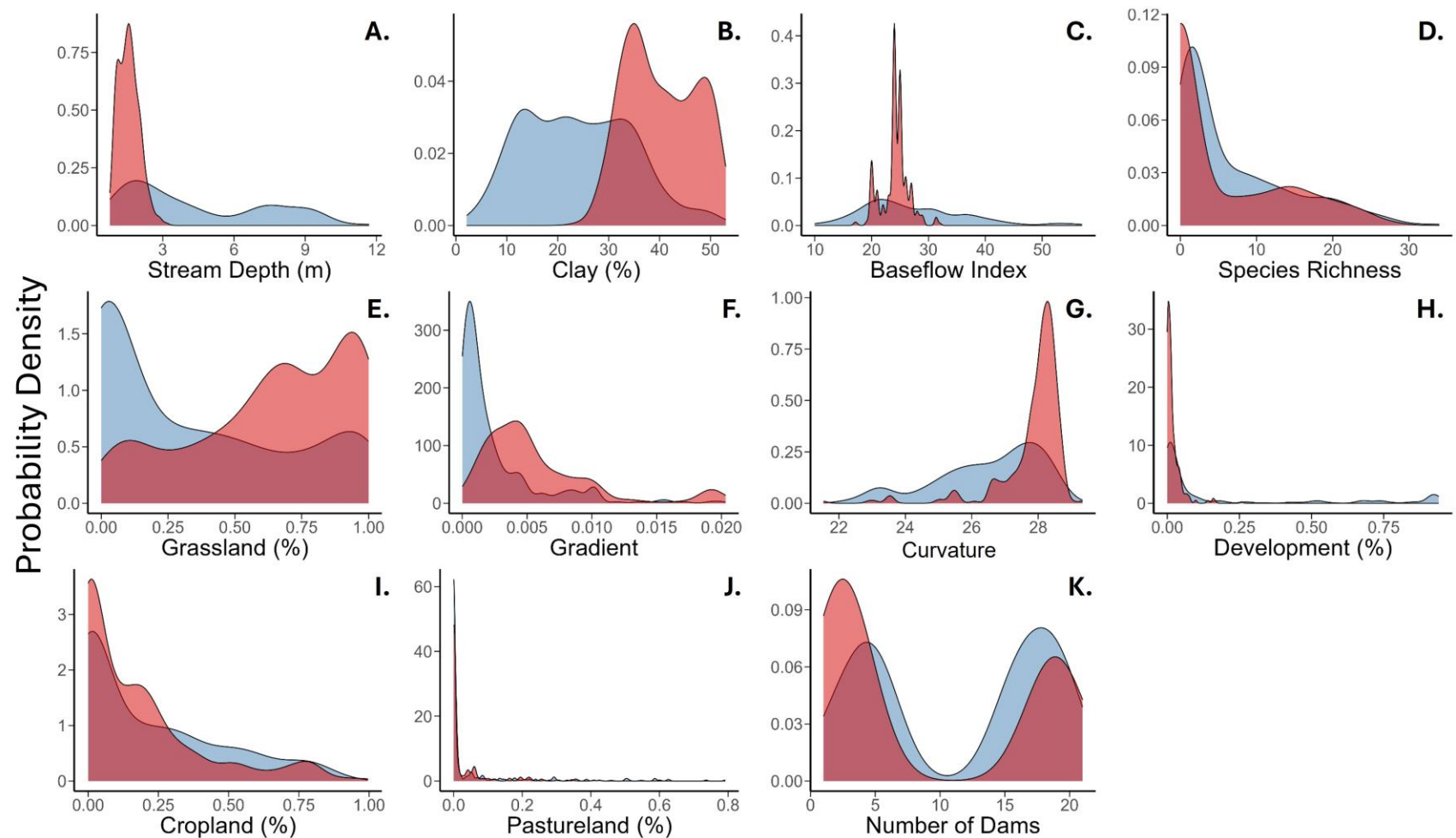


Figure 2.5. *Objective 2A – Step 9.* Multi-panel plot of *stream depth* (X axis) to evaluate consistency of trends in the Step 8 quantitative sub-steps (8A, 8B, 8F). The Y axis on panels represent (A) histogram frequencies (Sub-step 8B), (B) boxplots of Topeka Shiner presence ($n = 335$; 1) and absence (0) (Sub-step 8C), (C) KDE probability of fish presence (red) and absence (blue) (Sub-step 8C), (D) distributional habitat maps of Kansas, and predicted probability of Topeka Shiner presence from (E) single (Sub-step 8C), and (F) multiple logistic regression (Sub-step 8F). Data were collected from 670 stream sites in Kansas from 1947 to 2020.

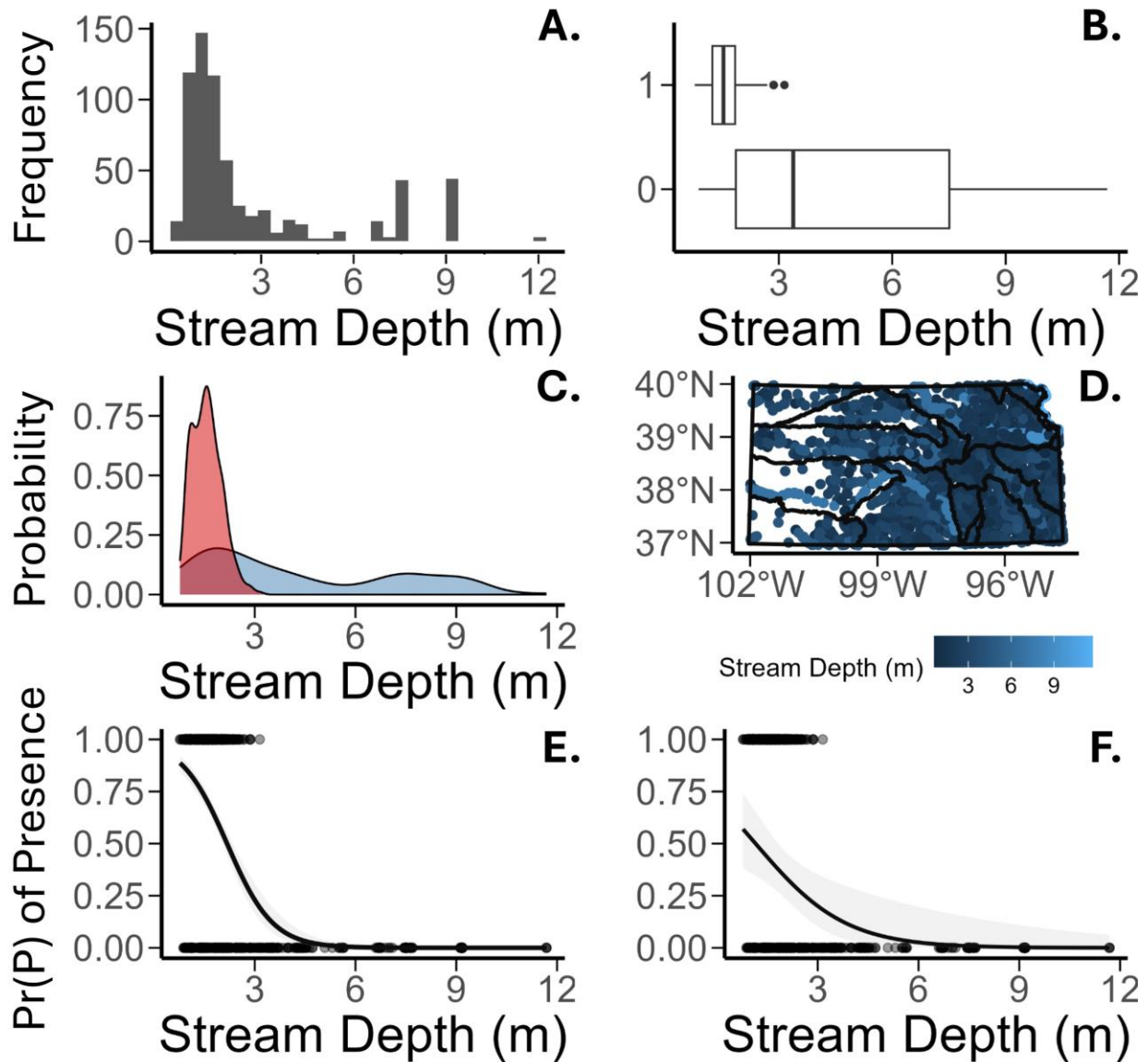


Figure 2.6. *Objective 2A – Step 9.* Multi-panel plot of *baseflow index* (X axis) to evaluate consistency of trends in the Step 8 quantitative sub-steps (8A, 8B, 8F). The Y axis on panels represent (A) histogram frequencies (Sub-step 8B), (B) boxplots of Topeka Shiner presence ($n = 335$; 1) and absence (0) (Sub-step 8C), (C) KDE probability of fish presence (red) and absence (blue) (Sub-step 8C), (D) distributional habitat maps of Kansas, and predicted probability of Topeka Shiner presence from (E) single (Sub-step 8C), and (F) multiple logistic regression (Sub-step 8F). Data were collected from 670 stream sites in Kansas from 1947 to 2020.

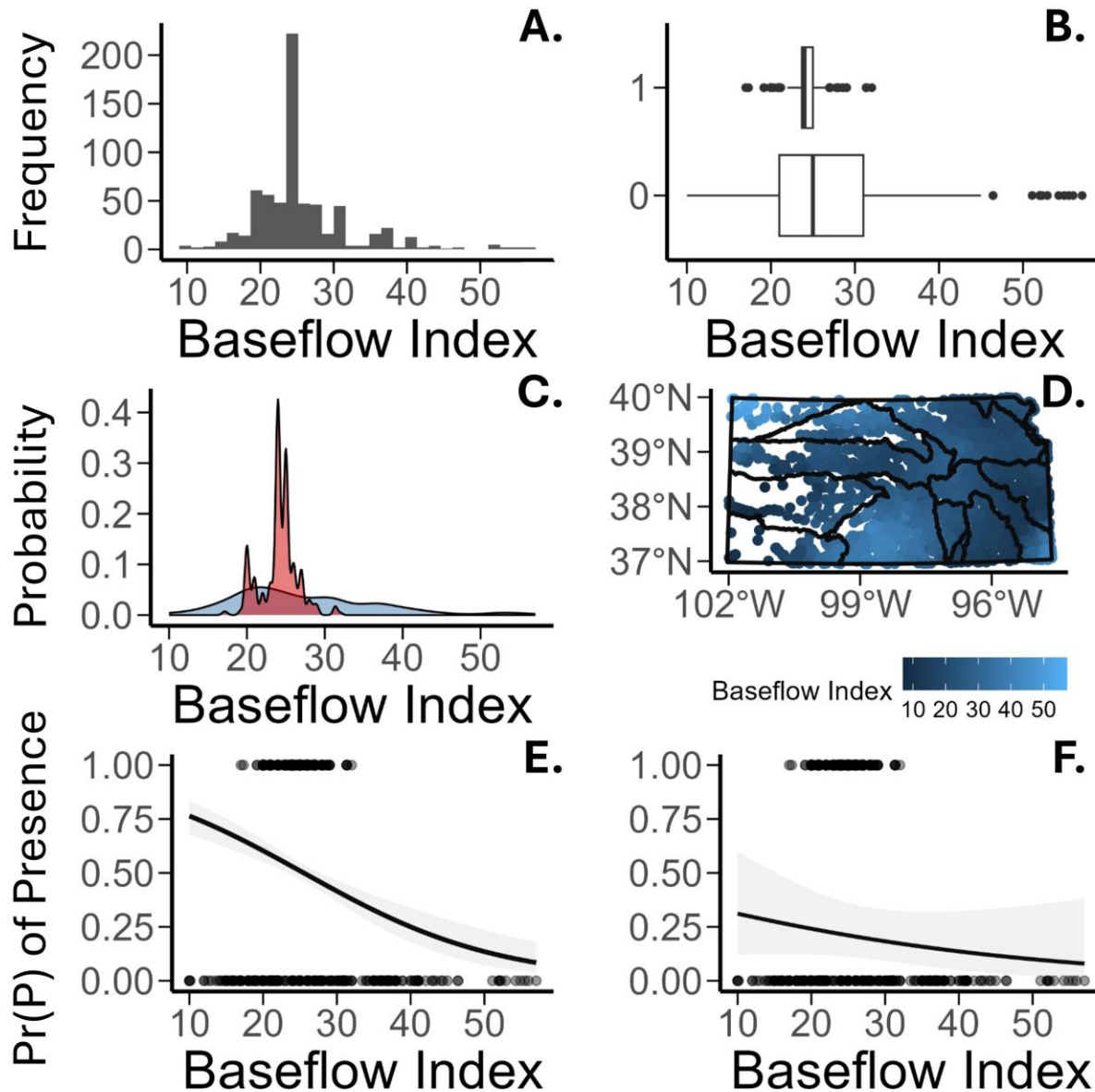


Figure 2.7. *Objective 2B – Step 8D.* Predicted probability of Topeka Shiner presence (Y axis) from single logistic regression models assessing the role of (A) a single linear parameter, (B) a single squared parameter, and (C) two quadratic parameters for the *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow) ecological variable (X axis). All three model estimations significantly ($p \leq 0.05$) describe Topeka Shiner presence ($n = 335$); however, the quadratic parameterization was most parsimonious ($\Delta AIC_c < 2$) and best describes the data pattern. Data were collected from 670 stream sites in Kansas from 1947 to 2020.

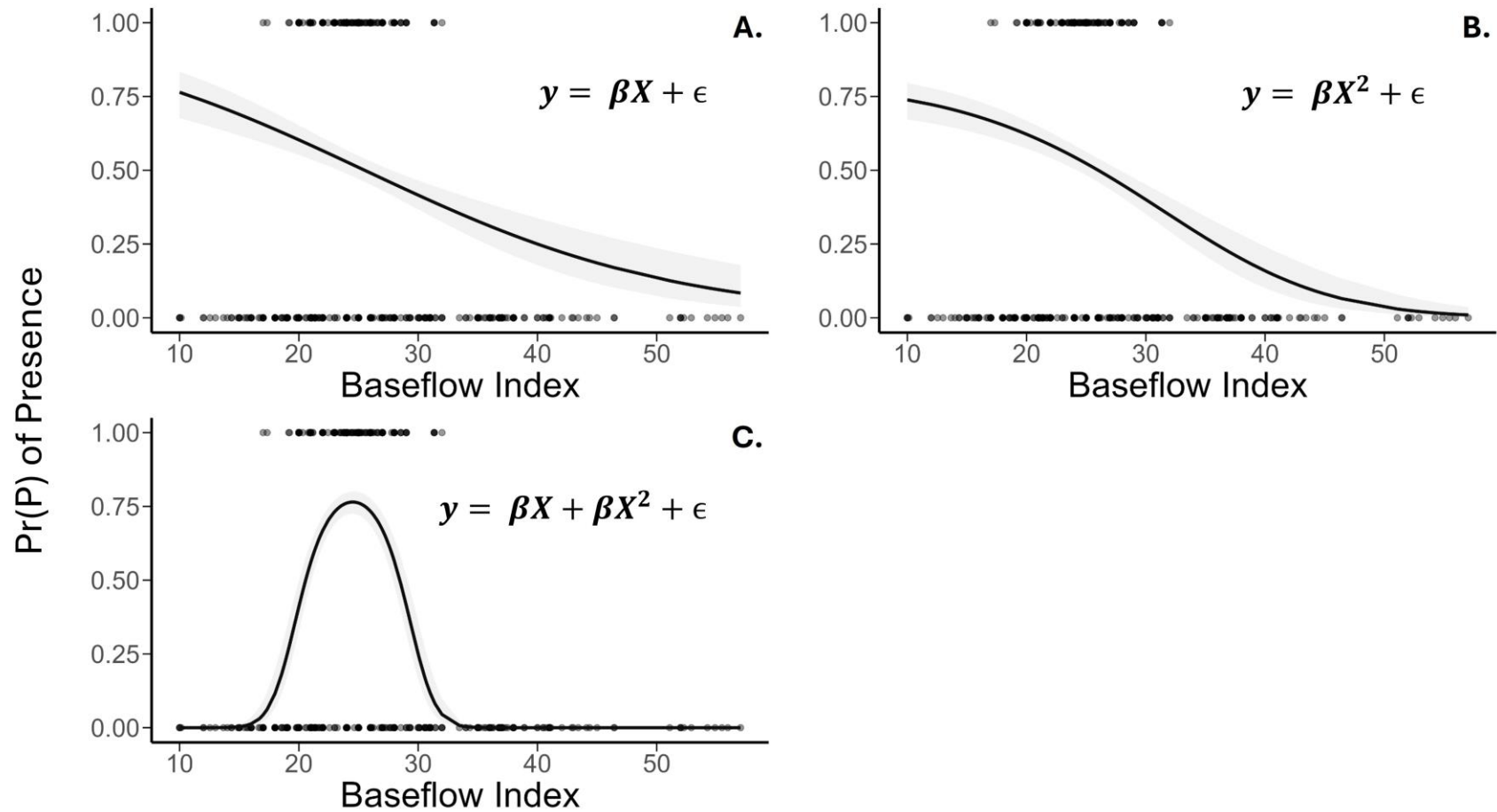


Figure 2.8. Objective 2B – Step 8F. Marginal effects on the predicted probability of Topeka Shiner presence (Y axis) from the top multiple logistic regression model that investigates the 11 ecological variables defined in Objective 1 plus *baseflow index*². Predictor ecological variables include: (A) *stream depth* (estimated in meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). Raw data distributions are depicted with black points at the top and bottom of each panel, response trends are black lines, and confidence intervals are shown in grey. The model results in this figure correspond to Table 3. Data were collected from 670 stream sites in Kansas from 1947 to 2020.

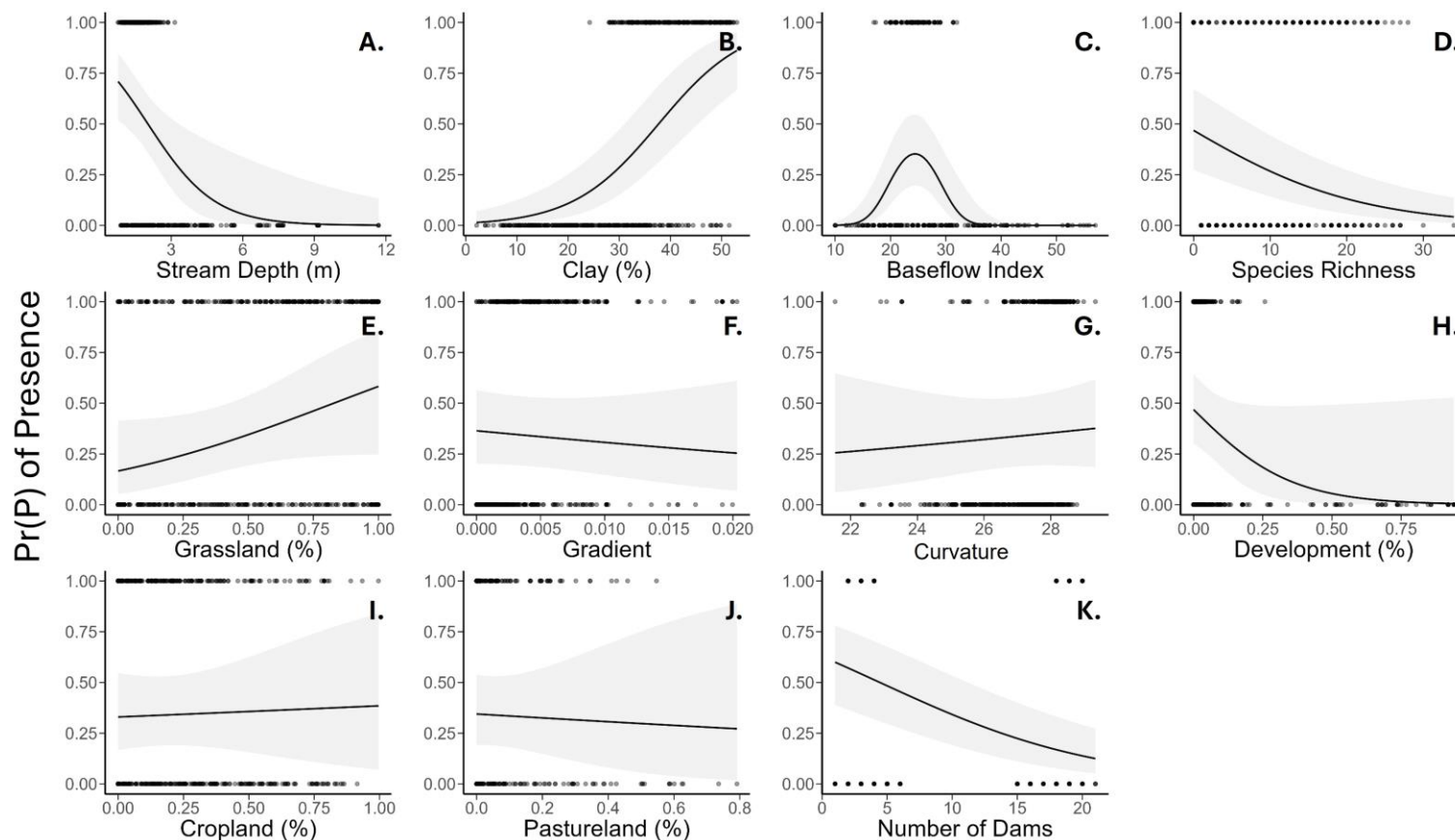


Figure 2.9. *Objective 2C – Step 8B-D.* Multi-panel plots with histogram sample frequencies (A-D), KDE probability ridge plots (E-H) with Topeka Shiner presence ($n = 335$; red) and absence (blue), boxplots of fish presence (1) and absence (0) (I-L), and predicted probability of fish presence from single logistic regression (M-P) on the Y axis. The four biotic variables on the X axes include *species richness* (number of fish species in a single sample), *predator species richness* (predators), *competitor species richness* (competitors), and *facilitator species richness* (facilitators) for communities sampled at 670 stream sites in Kansas between 1947 and 2020.

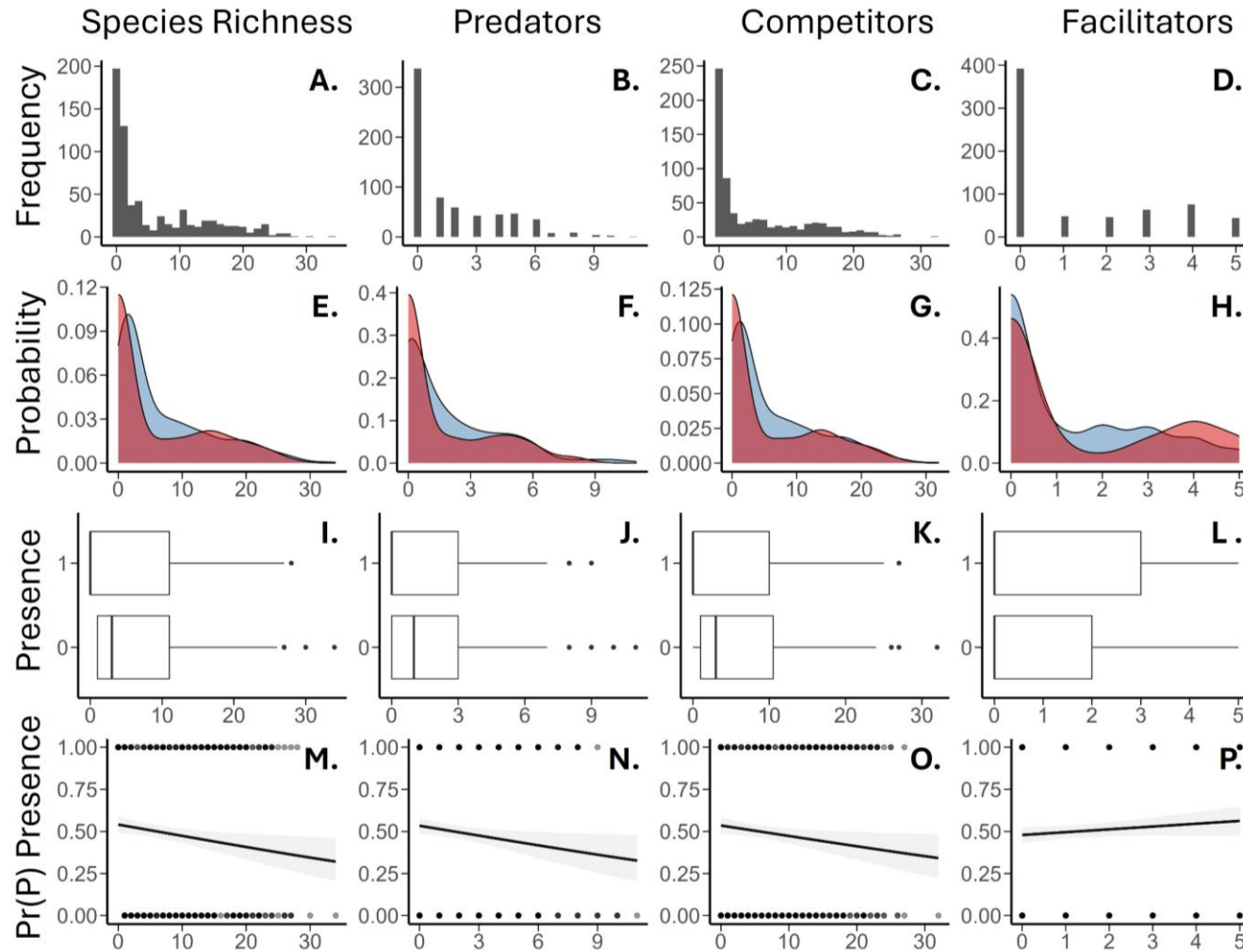


Figure 2.10. *Objective 2C – Step 8F.* Marginal effects on the predicted probability of Topeka Shiner presence (Y axis) from the top multiple logistic regression model for biotic investigation. Predictor ecological variables include: (A) *stream depth* (estimated in meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). Raw data distributions are depicted with black points at the top and bottom of each panel, response trends are black lines, and confidence intervals are shown in grey. The model results in this figure correspond to Table 2. Data were collected from 670 stream sites in Kansas from 1947 to 2020.

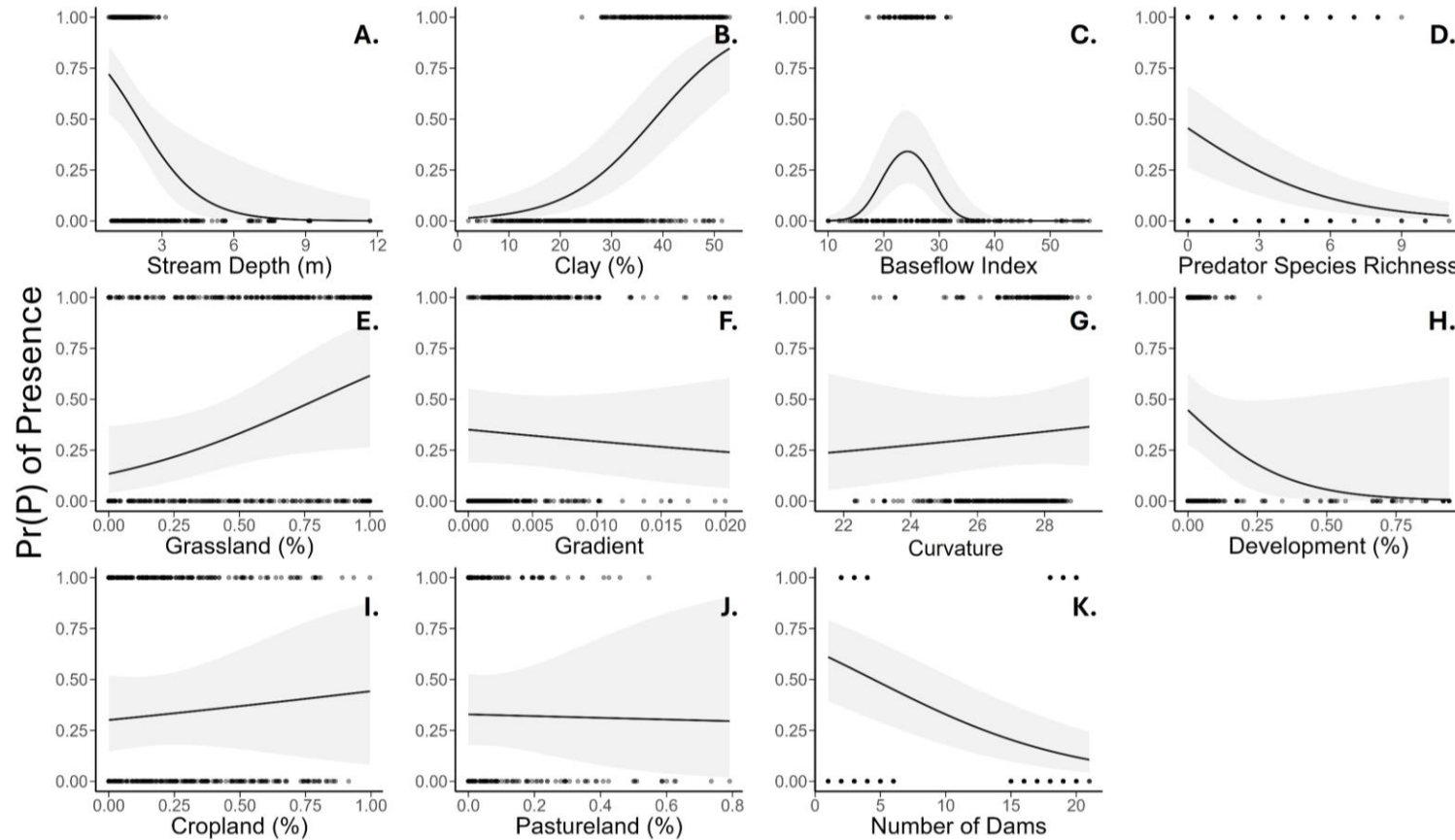


Figure 2.11. *Objective 2D – Step 8F.* Marginal effects on the predicted probability of Topeka Shiner presence (Y axis) from the interaction *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow) and *clay* (% in catchment surface geology). Predictor variables in the model include: (A) *stream depth* (estimated in meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). The model results in this figure correspond to Table 3. X axes show (A) *clay* and (B) *baseflow index*. Legends correspond to the multiple response curves as (A) *baseflow index* and (B) *clay* are held at specific levels. Data were collected from 670 stream sites in Kansas from 1947 to 2020.

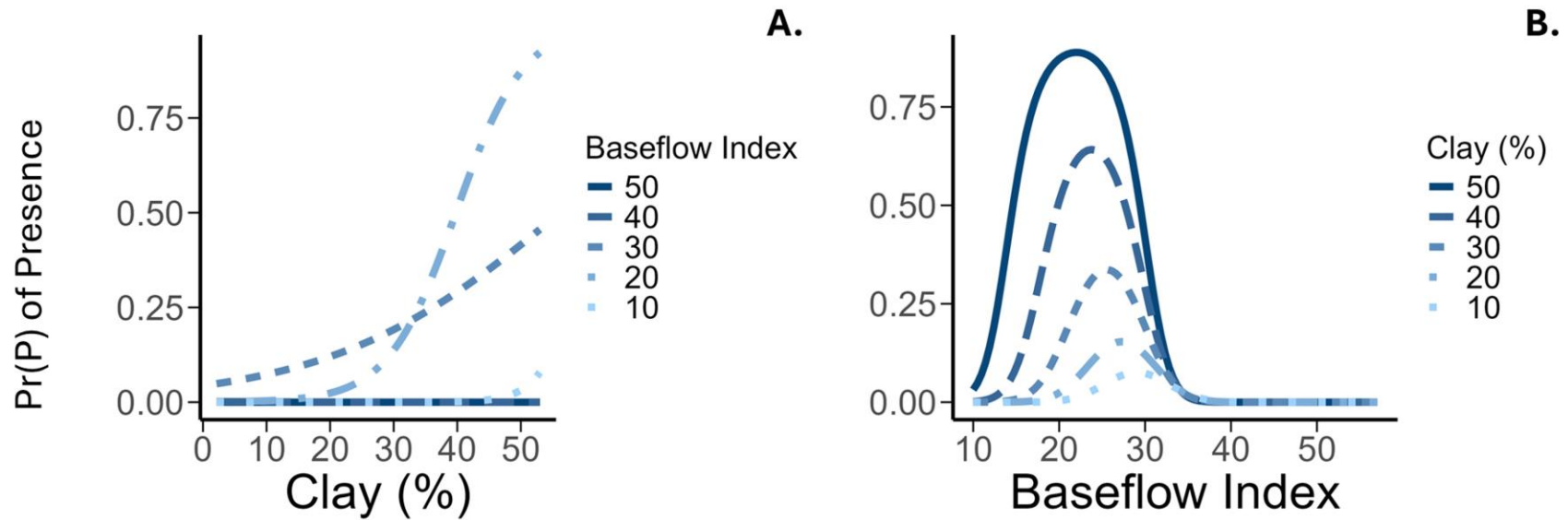


Figure 2.12. *Objective 2E – Step 8B.* (A) Map of major river basins (Basin) in Kansas (numbered 1-12). (B) Percentage (Y axis) of 670 stream sample sites sampled between 1947 to 2020 with Topeka Shiner presence ($n= 335$; red) and absence (blue) among the twelve major river basins (X axis).

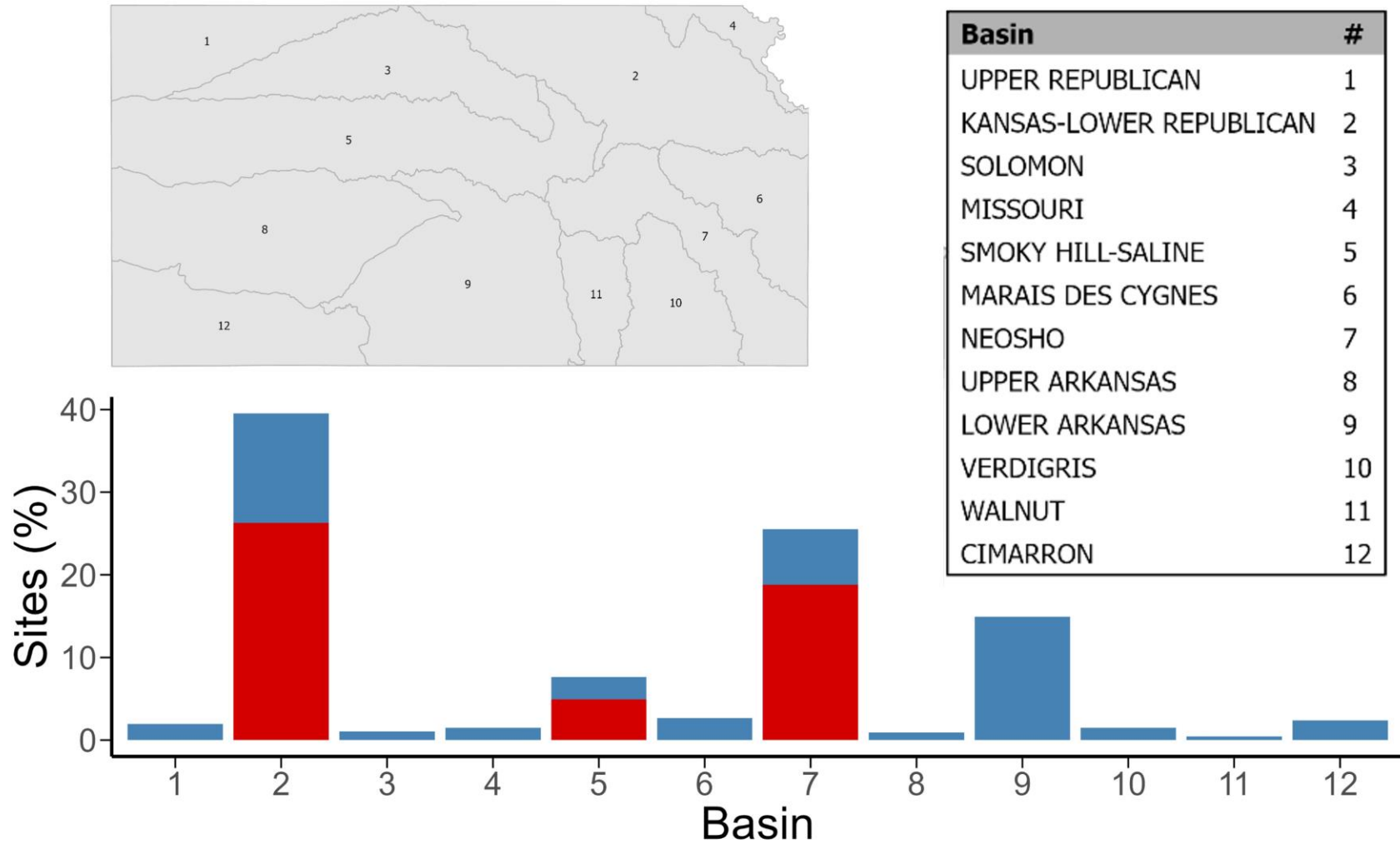
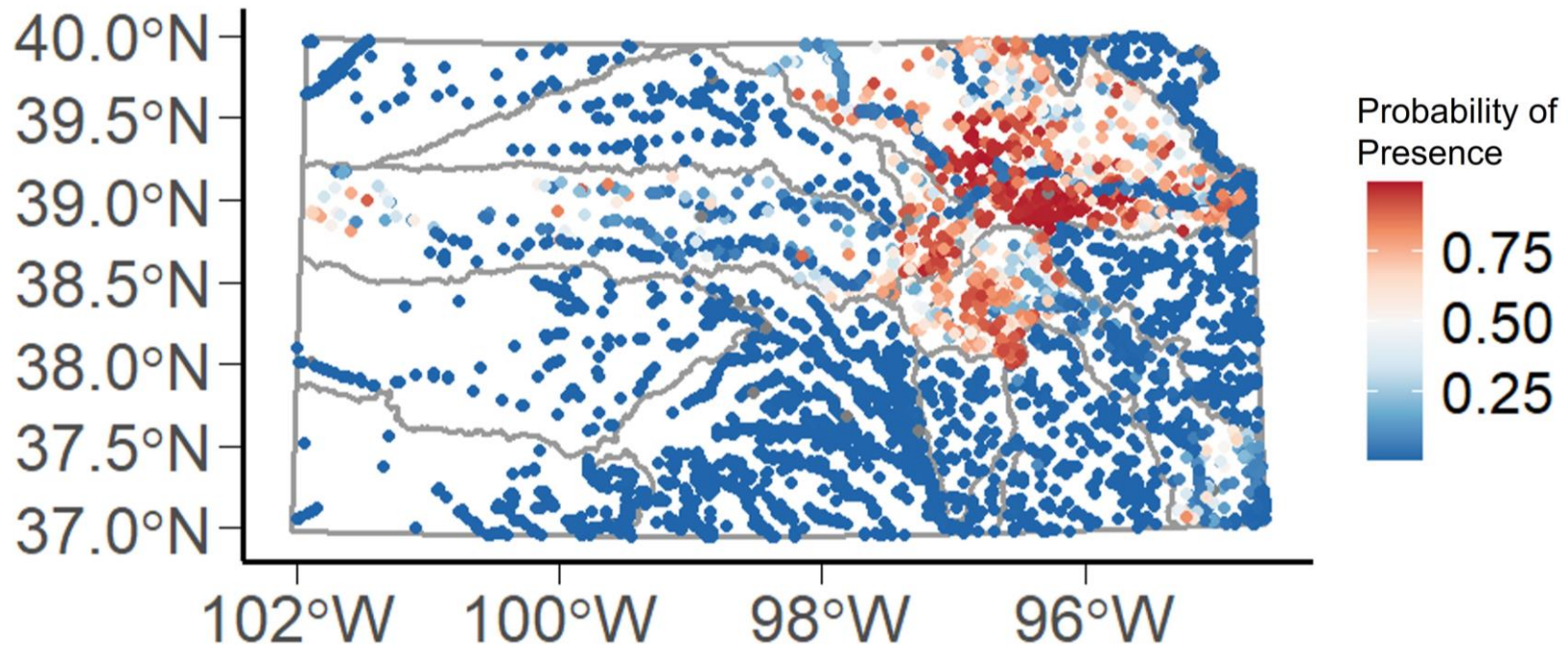


Figure 2.13. *Objective 2E – Step 8F.* Modeled predicted Topeka Shiner probability of presence (0 to 1; blue to red points) from mixed multiple logistic regression estimates (Objective 2E) for 12,086 stream sites in Kansas, USA, collected between 1857 and 2021. Eighty-one percent of data were collected between 2000-2021. Map axes correspond to latitude and longitude. The model results in this figure correspond to Table 4.



References

- Aadland, L. P., & Kuitunen, A. (n.d.). *Habitat suitability criteria for stream fishes and mussels of minnesota*.
- Adams, S., Hoover, J., & Killgore, K. (2000). Swimming performance of the Topeka shiner (*Notropis topeka*) an endangered midwestern minnow. *American Midland Naturalist*, 144(1), 178–186. (WOS:000088519700017). [https://doi.org/10.1674/0003-0031\(2000\)144%5B0178:SPOTT5D2.0.CO;2](https://doi.org/10.1674/0003-0031(2000)144%5B0178:SPOTT5D2.0.CO;2)
- Arnold, T. W. (2010). Uninformative Parameters and Model Selection Using Akaike's Information Criterion. *Journal of Wildlife Management*, 74(6), 1175–1178. <https://doi.org/10.2193/2009-367>
- Bakevich, B., Pierce, C., & Quist, M. (2015). Status of the Topeka Shiner in West-Central Iowa. *American Midland Naturalist*, 174(2), 350–358. (WOS:000361847900015).
- Barbarossa, V., Schmitt, R. J. P., Huijbregts, M. A. J., Zarfl, C., King, H., & Schipper, A. M. (2020). Impacts of current and future large dams on the geographic range connectivity of freshwater fish worldwide. *Proceedings of the National Academy of Sciences of the United States of America*, 117(7), 3648–3655. <https://doi.org/10.1073/pnas.1912776117>
- Bayless, M., McManus, M., & Fairchild, J. (2003). Geomorphic, water quality and fish community patterns associated with the distribution of *Notropis topeka* in a Central Missouri watershed. *American Midland Naturalist*, 150(1), 58–72. (WOS:000183990100006). [https://doi.org/10.1674/0003-0031\(2003\)150%5B0058:GWQAFC5D2.0.CO;2](https://doi.org/10.1674/0003-0031(2003)150%5B0058:GWQAFC5D2.0.CO;2)
- Becker, E. A., Carretta, J. V., Forney, K. A., Barlow, J., Brodie, S., Hoopes, R., Jacox, M. G., Maxwell, S. M., Redfern, J. V., Sisson, N. B., Welch, H., & Hazen, E. L. (2020).

- Performance evaluation of cetacean species distribution models developed using generalized additive models and boosted regression trees. *Ecology and Evolution*, 10(12), 5759–5784. <https://doi.org/https://doi.org/10.1002/ece3.6316>
- Berg, J. A., Petersen, T. A., Anderson, Y., & Baker, R. (2004). *Hydrogeology of the rock river watershed, minnesota, and associated off-channel habitats of the topeka shiner* [Final Report]. Minnesota Department of Natural Resources, Division of Waters and Ecological Services. <https://wrl.mnpals.net/islandora/object/WRLrepository%253A2760>
- Bieger, K., Rathjens, H., Allen, P. M., & Arnold, J. G. (2015). Development and Evaluation of Bankfull Hydraulic Geometry Relationships for the Physiographic Regions of the United States. *JAWRA Journal of the American Water Resources Association*, 51(3), 842–858. <https://doi.org/10.1111/jawr.12282>
- Bolker, B. M. (2008). *Ecological Models and Data in R*. Princeton University Press.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127–135. <https://doi.org/10.1016/j.tree.2008.10.008>
- Bopp, J., Robinson, K. F., Nathan, L., Herbst, S., Brenden, T. O., Mayer, C. M., & Dettmers, J. M. (2025). Reevaluation of an adaptive management framework for invasive Grass Carp within Lake Erie. *Transactions of the American Fisheries Society*, 154(5), 490–504. <https://doi.org/10.1093/tafafs/vnaf024>
- Bouska, W., & Paukert, C. (2010). Road crossing designs and their impact on fish assemblages of Great Plains streams. *Transactions of the American Fisheries Society*, 139(1), 214–222. (WOS:000277124800017). <https://doi.org/10.1577/T09-040.1>

- Bowler, D. E., Boyd, R. J., Callaghan, C. T., Robinson, R. A., Isaac, N. J. B., & Pocock, M. J. O. (2025). Treating gaps and biases in biodiversity data as a missing data problem. *Biological Reviews*, 100(1), 50–67. <https://doi.org/10.1111/brv.13127>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brooks, T., Fonseca, G. A. B. da, & Rodrigues, A. S. L. (2004). Species, data, and conservation planning. *Conservation Biology*, 18(6), 1682–1688.
- Campbell, S. W., Szuwalski, C. S., Tabor, V. M., & deNoyelles, F. (2016). Challenges to reintroduction of a captive population of Topeka Shiner (*Notropis topeka*) into former habitats in Kansas. *Transactions of the Kansas Academy of Science (1903-)*, 119(1), 83–92.
- Chollet Ramampandra, E., Scheidegger, A., Wydler, J., & Schuwirth, N. (2023). A comparison of machine learning and statistical species distribution models: Quantifying overfitting supports model interpretation. *Ecological Modelling*, 481, 110353. <https://doi.org/10.1016/j.ecolmodel.2023.110353>
- Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, 74(368), 829–836. <https://doi.org/10.1080/01621459.1979.10481038>
- Cook, C. N., Mascia, M. B., Schwartz, M. W., Possingham, H. P., & Fuller, R. A. (2013). Achieving conservation science that bridges the knowledge–action boundary. *Conservation Biology*, 27(4), 669–678. <https://doi.org/10.1111/cobi.12050>

- Cooper, A., & Infante, D. M. (2024). *Dam metrics representing stream fragmentation and flow alteration for the conterminous united states linked to the NHDPlusV2.1* [Dataset]. U.S. Geological Survey. <https://doi.org/10.5066/P94JQOFU>
- Cross, F. B., & Moss, R. E. (1987). Historic changes in fish communities and aquatic habitats in plains streams of Kansas. In D. C. Heins (Ed.), *Community and Evolutionary Ecology of North American Stream Fishes* (pp. 155–165). University of Oklahoma Press.
<https://hero.epa.gov/reference/3349146/>
- Cutler, D. R., Edwards Jr., T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random Forests for Classification in Ecology. *Ecology*, 88(11), 2783–2792.
<https://doi.org/10.1890/07-0539.1>
- Dewitz, J. (2023). *National Land Cover Database (NLCD) 2021 Products*.
<https://doi.org/10.5066/P9JZ7AO3>
- Dingman, S. L. (2015). *Physical hydrology*. Waveland press.
- Dodds, W. K., Gido, K., Whiles, M. R., Fritz, K. M., & Matthews, W. J. (2004). Life on the Edge: The Ecology of Great Plains Prairie Streams. *BioScience*, 54(3), Article 3.
[https://doi.org/10.1641/0006-3568\(2004\)054%5B0205:LOTETE%5D2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054%5B0205:LOTETE%5D2.0.CO;2)
- Dudley, P. N., John, S. N., Daniels, M. E., & Danner, E. M. (2022). Using decades of spawning data and hydraulic models to construct a temperature-dependent resource selection function for management of an endangered salmonid. *Canadian Journal of Fisheries and Aquatic Sciences*, 79(1), 73–81. <https://doi.org/10.1139/cjfas-2021-0022>
- Eberle, M., Edds, D., Gido, K., Haslouer, S., Huggins, D., Mosher, T., Stark, B., Tomelleri, J., Triplett, J., Wiley, E., & distler, d. (2014). *Kansas Fishes Committee (Distler, D., M. Eberle, D. Edds, K. Gido, S. Haslouer, D. Huggins, T. Mosher, B. Stark, J. Tomelleri, J.*

- Triplett, and E. Wiley.) 2014. *Kansas Fishes*. University Press of Kansas, Lawrence. 496 pages, 184 color illustrations, 121 maps, ISBN 978-0-7006-1961-0.
- Englund, G., & Cooper, S. D. (2003). *Scale effects and extrapolation in ecological experiments*.
[https://doi.org/10.1016/s0065-2504\(03\)33011-9](https://doi.org/10.1016/s0065-2504(03)33011-9)
- EPA, E. P. A. (2007, September). *Procedures for Implementing the National Environmental Policy Act and Assessing the Environmental Effects Abroad of EPA Actions* (Issue EPA-HQ-OECA-2005-0062; FRL-8467-5; RIN 2020-AA42). U.S. Environmental Protection Agency.
- Esri Inc. (2024). *ArcGIS Pro* (Version 3.5.4) [Computer software]. Esri Inc.
<https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>
- Faraway, J. J. (2016). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models* (2nd ed.). CRC Press.
- Fausch, K. D., Torgersen, C. E., Baxter, C. V., & Li, H. W. (2002). Landscapes to Riverscapes: Bridging the Gap between Research and Conservation of Stream Fishes: A Continuous View of the River is Needed to Understand How Processes Interacting among Scales Set the Context for Stream Fishes and Their Habitat. *BioScience*, 52(6), Article 6.
[https://doi.org/10.1641/0006-3568\(2002\)052%5B0483:LTRBTG%5D2.0.CO;2](https://doi.org/10.1641/0006-3568(2002)052%5B0483:LTRBTG%5D2.0.CO;2)
- Fischer, J., Bakevich, B., Shea, C., Pierce, C., & Quist, M. (2018). Floods, drying, habitat connectivity, and fish occupancy dynamics in restored and unrestored oxbows of West Central Iowa, USA. *Aquatic Conservation-Marine and Freshwater Ecosystems*, 28(3), 630–640. (WOS:000435930100012). <https://doi.org/10.1002/aqc.2896>
- Flather, C. (1997). *Identifying gaps in conservation networks: Of indicators and uncertainty in geographic-based analyses*. 7(2).

- Fletcher Jr., R. J., Green, R. E., Bladon, E. K., Atkinson, P. W., Phalan, B. T., Williams, D., Visconti, P., & Balmford, A. (2025). Beyond Species Richness for Biological Conservation. *Conservation Letters*, 18(4), e13124. <https://doi.org/10.1111/conl.13124>
- Frimpong, E. A., & Angermeier, P. L. (2009). Fish Traits: A Database of Ecological and Life-history Traits of Freshwater Fishes of the United States. *Fisheries*, 34(10), 487–495. <https://doi.org/10.1577/1548-8446-34.10.487>
- Galic, N., Schmolke, A., Bartell, S., Roy, C., & Brain, R. (2023). Applying a hybrid model to support management of the endangered Topeka shiner in oxbow habitats. *Frontiers in Ecology and Evolution*, 10. (WOS:000918515500001). <https://doi.org/10.3389/fevo.2022.1075244>
- Gerken, J. E., & Paukert, C. P. (2013). Fish assemblage and habitat factors associated with the distribution of Topeka shiner (*Notropis topeka*) in Kansas streams. *Journal of Freshwater Ecology*, 28(4), 503–516. <https://doi.org/10.1080/02705060.2013.792754>
- Gido, K. B., Dodds, W. K., & Eberle, M. E. (2010). Retrospective analysis of fish community change during a half-century of landuse and streamflow changes. *Journal of the North American Benthological Society*, 29(3), 970–987. <https://doi.org/10.1899/09-116.1>
- Gitzen, R. A., Millspaugh, J. J., Cooper, A. B., & Licht, D. S. (2012). *Design and Analysis of Long-Term Ecological Monitoring Studies*. Cambridge University Press. <http://ebookcentral.proquest.com/lib/ksu/detail.action?docID=944689>
- Gozlan, R. E., Britton, J. R., Cowx, I., & Copp, G. H. (2010). Current knowledge on non-native freshwater fish introductions. *Journal of Fish Biology*, 76(4), 751–786. <https://doi.org/10.1111/j.1095-8649.2010.02566.x>

- Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., ... Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*(12), 1424–1435.
<https://doi.org/10.1111/ele.12189>
- Haslouer, S. G., Eberle, M. E., Edds, D. R., Gido, K. B., Mammoliti, C. S., Triplett, J. R., Collins, J. T., Distler, D. A., Huggins, D. G., & Stark, W. J. (2005). Current status of native fish species in Kansas. *Transactions of the Kansas Academy of Science*, *108*(1 & amp; 2), Article 1 & amp; 2. [https://doi.org/10.1660/0022-8443\(2005\)108%5B0032:CSONFS%5D2.0.CO;2](https://doi.org/10.1660/0022-8443(2005)108%5B0032:CSONFS%5D2.0.CO;2)
- Hatch, J. (2001). What we know about Minnesota's first endangered fish species: The Topeka Shiner. *Journal of the Minnesota Academy of Science*, *65*(1), 39–46.
- Hazen, E. L., Abrahms, B., Brodie, S., Carroll, G., Welch, H., & Bograd, S. J. (2021). Where did they not go? Considerations for generating pseudo-absences for telemetry-based habitat models. *Movement Ecology*, *9*(1), 5. <https://doi.org/10.1186/s40462-021-00240-2>
- Hebert, N. E., Sameoto, J. A., Keith, D. M., Barrett, M. A., Brown, C. J., & Mills Flemming, J. (2024). Comparing spatiotemporal species distribution models: A case study of a Scotian Shelf sea cucumber (*Cucumaria frondosa*). *Ecosphere*, *15*(4), e4822.
<https://doi.org/10.1002/ecs2.4822>
- Hill, R. A., Weber, M. H., Leibowitz, S. G., Olsen, A. R., & Thornbrugh, D. J. (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. <https://doi.org/10.1111/1752-1688.12372>
- Hitchman, S. M., Mather, M. E., Smith, J. M., & Fencel, J. S. (2018). Habitat mosaics and path analysis can improve biological conservation of aquatic biodiversity in ecosystems with low-head dams. *Science of The Total Environment*, 619–620, 221–231. <https://doi.org/10.1016/j.scitotenv.2017.10.272>
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). John Wiley & Sons. <https://doi.org/10.1002/9781118548387>
- Huang, J., & Frimpong, E. A. (2015). Using Historical Atlas Data to Develop High-Resolution Distribution Models of Freshwater Fishes. *PLOS ONE*, 10(6), e0129995. <https://doi.org/10.1371/journal.pone.0129995>
- Hutto, R. L., & Belote, R. T. (2013). Distinguishing four types of monitoring based on the questions they address. *Forest Ecology and Management*, 289, 183–189. <https://doi.org/10.1016/j.foreco.2012.10.005>
- IDNR. (2015). *Iowa Wildlife Action Plan: Securing a future for fish and wildlife – a conservation legacy for Iowans*. Iowa Department of Natural Resources. <https://www.iowadnr.gov/programs-services/iowas-wildlife/wildlife-action-plan>
- Johnson, D. (1980). The Comparison of Usage and Availability Measurements for Evaluating Resource Preference. *United States Geological Survey, Northern Prairie Wildlife Research Center: Publications*. <https://digitalcommons.unl.edu/usgsnpwrc/198>
- Johnson, D. H. (2002). The importance of replication in wildlife research. *The Journal of Wildlife Management*, 66(4), 919–932. <https://doi.org/10.2307/3802926>

Kansas Legislature. (1917). *Permit or Consent of Chief Engineer Required to Construct, Modify or Add to Dams or Other Water Obstructions.*

https://www.ksrevisor.gov/statutes/chapters/ch82a/082a_003_0001.html

Kansas Legislature. (1984). *Kansas Water Appropriation Act.*

<https://www.ksrevisor.org/statutes/chapters/ch82a.html>

KDHE, K. D. of H. and E. (2020). *Stream Biological Monitoring Program: Quality Assurance Management Plan, Part III.* Kansas Department of Health and Environment, Bureau of Water. <https://www.coronavirus.kdheks.gov/DocumentCenter/View/11945/Stream-Biological-Monitoring-Program---Part-III-PDF>

KDWP. (2008). *Stream Monitoring and Assessment Program: RFM 2007 Landowner Survey Site Summaries.* Kansas Department of Wildlife and Parks, Environmental Services Section.

<https://www.ksoutdoors.gov/home/showpublisheddocument/1880/638967158897800000>

KDWP. (2022). *Kansas State Wildlife Action Plan.* Kansas Department of Wildlife and Parks. https://ksoutdoors.gov/content/download/47442/484423/version/3/file/Kansas+SWAP_Final+2022.pdf

KDWP, K. D. of W. & P. (2024, March 30). *Stream Survey and Monitoring Program.*

<https://ksoutdoors.com/Services/Publications/Fish-and-Wildlife-Management/Stream-Survey-and-Monitoring-Program>

KDWP, K. D. of W. & P., & USFWS, U. S. F. and W. S. (2023). *The Kansas Aquatic Species Conservation Agreement: A Programmatic Safe Harbor Agreement and Candidate Conservation Agreement with Assurances for Twenty-One Aquatic Species in Kansas.*

- Kansas Department of Wildlife and Parks. <https://kdwpt.ks.gov/Wildlife-Habitats/Wildlife-Conservation/Kansas-Aquatic-Species-Recovery-Program>
- Kerns, H., & Bonneau, J. (2002). Aspects of the life history and feeding habits of the topeka shiner (*notropis topeka*) in kansas. *Transactions of The Kansas Academy of Science*, 105, 125–142. [https://doi.org/10.1660/0022-8443\(2002\)105%5B0125:AOTLHA%5D2.0.CO;2](https://doi.org/10.1660/0022-8443(2002)105%5B0125:AOTLHA%5D2.0.CO;2)
- Knight, G. L., & Gido, K. B. (2005). Habitat use and susceptibility to predation of four prairie stream fishes: Implications for conservation of the endangered Topeka Shiner. *Copeia*, 2005(1), 38–47. <https://doi.org/10.1643/CE-04-226R1>
- Kuhn & Max. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>
- Kutner, M. H. (Ed.). (2005). *Applied linear statistical models* (5. ed). McGraw-Hill Irwin.
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18–22.
- Lindenmayer, D. B., Lavery, T., & Scheele, B. C. (2022). Why we need to invest in large-scale, long-term monitoring programs in landscape ecology and conservation biology. *Current Landscape Ecology Reports*, 7(4), 137–146. <https://doi.org/10.1007/s40823-022-00079-2>
- Lindenmayer, D. B., & Likens, G. E. (2010). The science and application of ecological monitoring. *Biological Conservation*, 143(6), Article 6. <https://doi.org/10.1016/j.biocon.2010.02.013>
- Long, J. A. (2024). *interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions*. <https://doi.org/10.32614/CRAN.package.interactions>

- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60), 3139. <https://doi.org/10.21105/joss.03139>
- Mammoliti, C. (2014). Topeka Shiner, *Notropis topeka* Gilbert 1884. *Kansas Fishes Committee. Kansas Fishes. University Press of Kansas, Lawrence*, 225–227.
- Mather, M., Kuck, S., & Oliver, D. (2026). Statistical Facilitation in Environmental Science: Integrating Results from Complementary Statistical Analyses Can Improve Ecological Interpretations. *Environments*, 13(2), 82. <https://doi.org/10.3390/environments13020082>
- Matthews, W. J. (1988). North American Prairie Streams as Systems for Ecological Study. *Journal of the North American Benthological Society*, 7(4), 387–409. <https://doi.org/10.2307/1467298>
- Mazerolle, M. J. (2023). *AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c)*. <https://cran.r-project.org/package=AICcmodavg>
- McGillcuddy, M., Warton, D. I., Popovic, G., & Bolker, B. M. (2025). Parsimoniously fitting large multivariate random effects in glmmTMB. *Journal of Statistical Software*, 112(1), 1–19. <https://doi.org/10.18637/jss.v112.i01>
- McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., & Rea, A. (2012). *NHDPlus Version 2: User Guide*. U.S. Environmental Protection Agency. <https://www.epa.gov/waterdata/nhdplus-national-hydrography-dataset-plus>
- Minckley, W. L., & Cross, F. B. (1959). Distribution, habitat, and abundance of the Topeka Shiner *Notropis topeka* (Gilbert) in Kansas. *The American Midland Naturalist*, 61(1), 210–217. <https://doi.org/10.2307/2422352>

- Mitchell, A. L., & Knouft, J. H. (2009). Non-native fishes and native species diversity in freshwater fish assemblages across the United States. *Biological Invasions*, 11(6), 1441–1450. <https://doi.org/10.1007/s10530-008-9352-9>
- MNDR. (2015). *Minnesota's Wildlife Action Plan 2015–2025*. Minnesota Department of Natural Resources. <https://files.dnr.state.mn.us/assistance/nrplanning/bigpicture/mnwap/wildlife-action-plan-2015-2025.pdf>
- Nielsen, S. E., Haughland, D. L., Bayne, E., & Schieck, J. (2009). Capacity of large-scale, long-term biodiversity monitoring programmes to detect trends in species prevalence. *Biodiversity and Conservation*, 18(11), 2961–2978. <https://doi.org/10.1007/s10531-009-9619-1>
- NRCS. (2022). *Web Soil Survey*. Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. <https://websoilsurvey.nrcs.usda.gov/>
- Olden, J. D., & Jackson, D. A. (2002). A comparison of statistical approaches for modelling fish species distributions. *Freshwater Biology*, 47(10), 1976–1995. <https://doi.org/10.1046/j.1365-2427.2002.00945.x>
- Olsen, A. R., Sedransk, J., Edwards, D., Gotway, C. A., Liggett, W., Rathbun, S., Reckhow, K. H., & Young, L. J. (1997). *Statistical Issues for Monitoring Ecological and Natural Resources in the United States*.
- O'Malley, W. C., Elbroch, L. M., Zeller, K. A., Beier, P., Beale, M. M., Beausoleil, R. A., Kertson, B., Knopff, K., Kunkel, K., Maletzke, B. T., Martins, Q., Matchett, M. R., Wilmers, C. C., Wittmer, H. U., Vickers, W., Sager-Fradkin, K., & Robinson, H. (2024). Machine learning allows for large-scale habitat prediction of a wide-ranging carnivore

- across diverse ecoregions. *Landscape Ecology*, 39(5), 106.
<https://doi.org/10.1007/s10980-024-01903-2>
- Ommer, R. E., & Perry, R. I. (2022). The importance of complexities of scale and context in fisheries research and management. *ICES Journal of Marine Science*, 79(8), 2196–2203.
<https://doi.org/10.1093/icesjms/fsac156>
- Ostroff, A., Wieferich, D., Cooper, A., & Infante, D. (2013). *2012 National Anthropogenic Barrier Dataset (NABD)*. U. S. Geological Survey Aquatic GAP Program.
<https://www.sciencebase.gov/catalog/item/56a7f9dce4b0b28f1184dabd>
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1), 439–446. <https://doi.org/10.32614/RJ-2018-009>
- Perkin, J. S., Gido, K. B., Costigan, K. H., Daniels, M. D., & Johnson, E. R. (2015). Fragmentation and drying ratchet down Great Plains stream fish diversity. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 25(5), Article 5.
<https://doi.org/10.1002/aqc.2501>
- Pfaff, P., Hase, K., & Gido, K. (2023). Community assembly of prairie farm ponds: Build it and they will come, stock it and they won't. *Canadian Journal of Fisheries and Aquatic Sciences*. (WOS:000910102300001). <https://doi.org/10.1139/cjfas-2022-0075>
- Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegard, K. L., Richter, B. D., Sparks, R. E., & Stromberg, J. C. (1997). The Natural Flow Regime. *BioScience*, 47(11), 769–784.
<https://doi.org/10.2307/1313099>
- Potter, K. M., Koch, F. H., Oswalt, C. M., & Iannone, B. V. (2016). Data, data everywhere: Detecting spatial patterns in fine-scale ecological information collected across a continent. *Landscape Ecology*, 31(1), 67–84. <https://doi.org/10.1007/s10980-015-0295-0>

- Price, K. (2011a). Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review. *Progress in Physical Geography*, 35(4), 465–492. <https://doi.org/10.1177/0309133311402714>
- Price, K. (2011b). Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review. *Progress in Physical Geography*, 35(4), 465–492. <https://doi.org/10.1177/0309133311402714>
- R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Radinger, J., Britton, J. R., Carlson, S. M., Magurran, A. E., Alcaraz-Hernández, J. D., Almodóvar, A., Benejam, L., Fernández-Delgado, C., Nicola, G. G., Oliva-Paterna, F. J., Torralva, M., & García-Berthou, E. (2019). Effective monitoring of freshwater fish. *Fish and Fisheries*, 20(4), Article 4. <https://doi.org/10.1111/faf.12373>
- Rahel, F. J., & Lynch, A. J. (2022). Managing fisheries within a RAD framework: Concepts and applications. *Fisheries Management and Ecology*, 29(4), 323–328. <https://doi.org/10.1111/fme.12578>
- Ranganathan, P., Pramesh, C. S., & Buyse, M. (2016). Common pitfalls in statistical analysis: The perils of multiple testing. *Perspectives in Clinical Research*, 7(2), 106. <https://doi.org/10.4103/2229-3485.179436>
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinformatics*, 12, 77. <https://doi.org/10.1186/1471-2105-12-77>
- Rode, O., Mather, M., Oliver, D., Nelson, K., Reed, V., Moore, T., & Pratap, S. (2025). A framework tool that applies weight-of-evidence integration to the analysis of existing

- datasets to guide freshwater conservation. *Frontiers in Freshwater Science*, 3.
<https://doi.org/10.3389/ffwsc.2025.1520312>
- Rohweder, M. R. (2022, June). *Kansas Wildlife Action Plan*. Ecological Services Section, Kansas Department of Wildlife and Parks in cooperation with the Kansas Biological Survey. 183 pp. 3rd Edition. <https://ksoutdoors.com/Services/Kansas-SWAP>
- Rosenfeld, J. S. (2017). Developing flow–ecology relationships: Implications of nonlinear biological responses for water management. *Freshwater Biology*, 62(8), 1305–1324.
<https://doi.org/10.1111/fwb.12948>
- Salafsky, N., Boshoven, J., Burivalova, Z., Dubois, N. S., Gomez, A., Johnson, A., Lee, A., Margoluis, R., Morrison, J., Muir, M., Pratt, S. C., Pullin, A. S., Salzer, D., Stewart, A., Sutherland, W. J., & Wordley, C. F. R. (2019). Defining and using evidence in conservation practice. *Conservation Science and Practice*, 1(5), e27.
<https://doi.org/10.1111/csp2.27>
- Schmolke, A., Bartell, S. M., Roy, C., Green, N., Galic, N., & Brain, R. (2019). Species-specific population dynamics and their link to an aquatic food web: A hybrid modeling approach. In *Ecological Modelling* (Vol. 405, pp. 1–14). Elsevier Science BV.
<https://doi.org/10.1016/j.ecolmodel.2019.03.024>
- Schneider, R., Stoner, K., Steinauer, G., Panella, M. J., & Humpert, M. (Eds.). (2011). *The Nebraska Natural Legacy Project: State Wildlife Action Plan* (2nd ed.). Nebraska Game and Parks Commission. <https://govdocs.nebraska.gov/epubs/G1000/B184-2011.pdf>
- Schrank, S., Guy, C., Whiles, M., & Brock, B. (2001). Influence of instream and landscape-level factors on the distribution of Topeka shiners *Notropis topeka* in Kansas streams. *Copeia*,

- (2), 413–421. (WOS:000168663700010). [https://doi.org/10.1643/0045-8511\(2001\)001%5B0413:IOIALL%5D2.0.CO;2](https://doi.org/10.1643/0045-8511(2001)001%5B0413:IOIALL%5D2.0.CO;2)
- SDDGFP. (2023). *Appendices: South Dakota Source Water Assessment and Protection Program*. South Dakota Department of Game, Fish and Parks. https://gfp.sd.gov/UserDocs/nav/15_Appendices_SDWAP.pdf
- Shoemaker, K. T., Heffelfinger, L. J., Jackson, N. J., Blum, M. E., Wasley, T., & Stewart, K. M. (2018). A machine-learning approach for extending classical wildlife resource selection analyses. *Ecology and Evolution*, 8(6), 3556–3569. <https://doi.org/10.1002/ece3.3936>
- Simpson, N., Bybel, A. P., Weber, M. J., Pierce, C. L., & Roe, K. J. (2019). Occurrence, abundance and associations of Topeka shiners () in restored and unrestored oxbows in Iowa and Minnesota, USA. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 29(10), 1735–1748. <https://doi.org/10.1002/aqc.3186>
- Simpson, N., Bybel, A., Weber, M., Pierce, C., & Roe, K. (2021). Factors associated with distributions of six fishes of greatest conservation need in streams in midwestern USA. *Aquatic Conservation-Marine and Freshwater Ecosystems*, 31(7), 1831–1846. (WOS:000657326300001). <https://doi.org/10.1002/aqc.3562>
- Stark, W. J., Luginbill, J. S., & Eberle, M. E. (2002). Natural History of a Relict Population of Topeka Shiner (*Notropis topeka*) in Northwestern Kansas. *Transactions of the Kansas Academy of Science*, 105(3), 143–152. [https://doi.org/10.1660/0022-8443\(2002\)105%5B0143:NHOARP%5D2.0.CO;2](https://doi.org/10.1660/0022-8443(2002)105%5B0143:NHOARP%5D2.0.CO;2)
- Steel, E. A., Hughes, R. M., Fullerton, A. H., Schmutz, S., Young, J. A., Fukushima, M., Muhar, S., Poppe, M., Feist, B. E., & Trautwein, C. (2010). Are We Meeting the Challenges of

- Landscape-Scale Riverine Research? A Review. *Living Reviews in Landscape Research*, 4(1), 1. <https://doi.org/10.12942/lrlr-2010-1>
- Steen, V. A., Tingley, M. W., Paton, P. W. C., & Elphick, C. S. (2021). Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution*, 12(2), 216–226. <https://doi.org/10.1111/2041-210X.13525>
- Strang, G., & Herman, E. “Jed.” (2016a, March 30). 3.2 *The Derivative as a Function—Calculus Volume 1* | OpenStax. OpenStax. <https://openstax.org/books/calculus-volume-1/pages/3-2-the-derivative-as-a-function>
- Strang, G., & Herman, E. “Jed.” (2016b, March 30). 4.3 *Maxima and Minima—Calculus Volume 1* | OpenStax. OpenStax. <https://openstax.org/books/calculus-volume-1/pages/4-3-maxima-and-minima>
- Sutherland, C., Hare, D., Johnson, P. J., Linden, D. W., Montgomery, R. A., & Droge, E. (2023). Practical advice on variable selection and reporting using Akaike information criterion. *Proceedings of the Royal Society B: Biological Sciences*, 290(2007), 20231261. <https://doi.org/10.1098/rspb.2023.1261>
- USFWS. (1997, October 24). *Endangered and Threatened Wildlife and Plants; Proposed Rule To List the Topeka Shiner as Endangered*. <https://www.federalregister.gov/documents/1997/10/24/97-28231/endangered-and-threatened-wildlife-and-plants-proposed-rule-to-list-the-topeka-shiner-as-endangered>
- USFWS. (1998). Endangered and Threatened Wildlife and Plants: Final Rule to List the Topeka Shiner as Endangered. *Endangered Species Bulletin*. <https://digitalcommons.unl.edu/endangeredspeciesbull/38>

- USFWS, U. S. F. and W. S. (2024). *Recovery Implementation Strategy for the Topeka Shiner (Notropis topeka)*. U.S. Fish and Wildlife Service.
https://ecos.fws.gov/docs/recovery_plan/20240813_Topeka_Shiner_RIS_v1.pdf
- USFWS, U. S. F. and W. S., & NMFS, N. M. F. S. (1986, June). *Interagency Cooperation—Endangered Species Act of 1973, as Amended*. U.S. Department of the Interior and U.S. Department of Commerce.
- Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., & Cushing, C. E. (1980). The River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(1), 130–137. <https://doi.org/10.1139/f80-017>
- Wahl, C., Galic, N., Brain, R., Vaugeois, M., Weber, M., Roe, K., Stewart, T., Utrup, N., Mauldin, L., Kenney, A., & Terui, A. (2024). An approximation approach to spatial connectivity for a data-limited endangered species with implications for habitat restoration. *Biological Conservation*, 291. (WOS:001185894100001).
<https://doi.org/10.1016/j.biocon.2024.110470>
- Wall, S., & Berry, C. (2004). Threatened fishes of the world: Notropis topeka Gilbert, 1884 (Cyprinidae). *Environmental Biology of Fishes*, 70(3), 246–246.
(WOS:000222345600007). <https://doi.org/10.1023/B:EBFI.0000033348.37110.de>
- Wall, S., Berry, C., Blausey, C., Jenks, J., & Kopplin, C. (2004). Fish-habitat modeling for gap analysis to conserve the endangered Topeka shiner (Notropis topeka). *Canadian Journal of Fisheries and Aquatic Sciences*, 61(6), 954–973. (WOS:000223096000009).
<https://doi.org/10.1139/F04-017>
- Ward, J. V., Stanford, J. A., Ward, J. V., & Stanford, J. A. (1983). The serial discontinuity concept of lotic ecosystems. *Fontaine, T D Iii And S M Bartell (Ed) Dynamics Of Lotic*

Ecosystems Xiii+494p Ann Arbor Science Publishers Ann Arbor, Mich , Usa Illus Issbn,
(Ed), Article Ed.

Waterman, B. R., Alcantar, G., Thomas, S. G., & Kirk, M. F. (2022). Spatiotemporal variation in runoff and baseflow in watersheds located across a regional precipitation gradient.

Journal of Hydrology: Regional Studies, 41, 101071.

<https://doi.org/10.1016/j.ejrh.2022.101071>

Webb, J. A., Arthington, A. H., & Olden, J. D. (2017). Chapter 14—Models of Ecological Responses to Flow Regime Change to Inform Environmental Flows Assessments. In A. C. Horne, J. A. Webb, M. J. Stewardson, B. Richter, & M. Acreman (Eds.), *Water for the Environment* (pp. 287–316). Academic Press.

<https://doi.org/https://doi.org/10.1016/B978-0-12-803907-6.00014-0>

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

<https://ggplot2.tidyverse.org>

Williams, B. K. (2011). Adaptive management of natural resources—Framework and issues.

Journal of Environmental Management, Adaptive Management for Natural Resources, 92(5), Article 5. <https://doi.org/10.1016/j.jenvman.2010.10.041>

Winston, M. (2002). Spatial and Temporal Species Associations with the Topeka shiner

(*Notropis topeka*) in Missouri. *Journal of Freshwater Ecology*, 17(2), 249–261.

(WOS:000175646300010). <https://doi.org/10.1080/02705060.2002.9663893>

Winter, T. C., Harvey, J. W., Franke, O. L., & Alley, W. M. (1998). Ground water and surface water: A single resource. In *Circular* (No. 1139). U.S. Geological Survey.

<https://doi.org/10.3133/cir1139>

Wishart, D. J. (2004). *Encyclopedia of the Great Plains*. University of Nebraska Press.

- Witte, C., Wildhaber, M., Arab, A., & Noltie, D. (2009). Substrate choice of territorial male Topeka shiners (*Notropis topeka*) in the absence of sunfish (*Lepomis* sp.). *Ecology of Freshwater Fish*, 18(3), 350–359. (WOS:000268924900002).
<https://doi.org/10.1111/j.1600-0633.2008.00351.x>
- Yates, A. G., & Bailey, R. C. (2006). The Stream and Its Altered Valley: Integrating Landscape Ecology into Environmental Assessments of Agro-Ecosystems. *Environmental Monitoring and Assessment*, 114(1), 257–271. <https://doi.org/10.1007/s10661-006-4779-x>
- Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. Springer. <https://doi.org/10.1007/978-0-387-87458-6>

Appendix A - Chapter 1

Tables

Appendix A Table 1. Estimates from multiple logistic regression on original data using 11 ecological explanatory variables to model the response of Topeka Shiner (*Notropis topeka*) probability of presence in Kansas from 1947 to 2020 ($n=670$). Significant estimates ($p \leq 0.05$) are denoted with an *.

Variable	β	SE	Z	95% CI		p		OR	VIF
				Lower	Upper				
(Intercept)	-4.84	3.94	-1.23	-12.56	2.88	0.22		0.01	
<i>Clay</i>	0.17	0.02	7.84	0.13	0.21	0.00	*	1.19	1.30
<i>Stream Depth</i>	-0.75	0.27	-2.80	-1.27	-0.22	0.01	*	0.47	1.77
<i>Number of Dams</i>	-0.10	0.02	-4.84	-0.15	-0.06	0.00	*	0.90	1.78
<i>Species Richness</i>	-0.09	0.02	-4.87	-0.13	-0.05	0.00	*	0.91	1.28
<i>Development</i>	-4.65	2.83	-1.64	-10.18	0.89	0.10		0.01	1.38
<i>Gradient</i>	6.34	38.03	0.17	-68.20	80.87	0.87		564.36	2.16
<i>Pastureland</i>	-2.24	1.97	-1.13	-6.10	1.63	0.26		0.11	2.84
<i>Baseflow Index</i>	-0.04	0.03	-1.18	-0.09	0.02	0.24		0.97	1.30
<i>Grassland</i>	2.00	1.15	1.75	-0.24	4.25	0.08		7.41	9.91
<i>Cropland</i>	0.38	1.22	0.31	-2.01	2.77	0.75		1.46	6.31
<i>Curvature</i>	0.08	0.13	0.62	-0.17	0.33	0.54		1.08	1.17

Appendix A Table 2. Summary statistics (mean $\pm$ standard deviation (SD)) for variables included in models of Topeka Shiner (*Notropis topeka*) probability of presence in Kansas from 1947 to 2020 ($n=670$).

Variable	Mean	SD
<i>Baseflow Index</i>	25.69	6.80
<i>Clay</i>	32.14	12.06
<i>Cropland</i>	0.20	0.24
<i>Development</i>	0.10	0.23
<i>Gradient</i>	0.00	0.00
<i>Grassland</i>	0.49	0.36
<i>Number of Dams</i>	10.07	7.61
<i>Pastureland</i>	0.04	0.11
<i>Curvature</i>	27.19	1.47
<i>Species Richness</i>	6.06	7.55
<i>Stream Depth</i>	2.95	2.49

Appendix B - Chapter 2

Tables

Appendix B Table 1. List of fish species recorded in the Kansas Stream Assessment and Monitoring dataset used in the calculation of species richness in analyses. Included are common name, scientific name, taxonomic family, and assigned trait guild (if applicable). Species are listed with trait guilds of potential competitors (with invertivore diets), predators (large piscivores), and facilitators (nest building and egg guarding) of Topeka Shiner as assigned by the FishTraits Database and expert advice. Species that are not assigned to a guild are left blank (--).

Common name	Scientific name	Family	Trait Guilds
Arkansas Darter	<i>Etheostoma cragini</i>	Percidae	--
Arkansas River Shiner	<i>Notropis girardi</i>	Cyprinidae	--
Banded Darter	<i>Etheostoma zonale</i>	Percidae	--
Banded Sculpin	<i>Cottus carolinae</i>	Cottidae	Competitor
Bigeye Shiner	<i>Notropis boops</i>	Cyprinidae	Competitor
Bighead Carp	<i>Hypophthalmichthys nobilis</i>	Cyprinidae	Competitor
Bigmouth Buffalo	<i>Ictiobus cyprinellus</i>	Catostomidae	Competitor
Bigmouth Shiner	<i>Notropis dorsalis</i>	Cyprinidae	Competitor
Black Buffalo	<i>Ictiobus niger</i>	Catostomidae	Competitor
Black Bullhead	<i>Ameiurus melas</i>	Ictaluridae	Predator, Competitor
Black Carp	<i>Mylopharyngodon piceus</i>	Cyprinidae	--
Black Crappie	<i>Pomoxis nigromaculatus</i>	Centrarchidae	Predator, Competitor
Black Redhorse	<i>Moxostoma duquesnei</i>	Catostomidae	Competitor
Blackside Darter	<i>Percina maculata</i>	Percidae	Competitor
Blackstripe Topminnow	<i>Fundulus notatus</i>	Fundulidae	Competitor
Blue Catfish	<i>Ictalurus furcatus</i>	Ictaluridae	Predator, Competitor
Blue Sucker	<i>Cycleptus elongatus</i>	Catostomidae	Competitor
Bluegill	<i>Lepomis macrochirus</i>	Centrarchidae	Competitor
Bluntnose Shiner	<i>Cyprinella camura</i>	Cyprinidae	Competitor
Bluntnose Darter	<i>Etheostoma chlorosomum</i>	Percidae	Competitor
Bluntnose Minnow	<i>Pimephales notatus</i>	Cyprinidae	Competitor
Brassy Minnow	<i>Hybognathus hankinsoni</i>	Cyprinidae	Competitor
Brindled Madtom	<i>Noturus miurus</i>	Ictaluridae	Competitor
Brook Silverside	<i>Labidesthes sicculus</i>	Atherinopsidae	Competitor

Appendix B Table 1. List of fish species recorded in the Kansas Stream Assessment and Monitoring dataset used in the calculation of species richness in analyses. Included are common name, scientific name, taxonomic family, and assigned trait guild (if applicable). Species are listed with trait guilds of potential competitors (with invertivore diets), predators (large piscivores), and facilitators (nest building and egg guarding) of Topeka Shiner as assigned by the FishTraits Database and expert advice. Species that are not assigned to a guild are left blank (--).

Common name	Scientific name	Family	Trait Guilds
Brown Trout	<i>Salmo trutta</i>	Salmonidae	Competitor
Bullhead Minnow	<i>Pimephales vigilax</i>	Cyprinidae	Competitor
Cardinal Shiner	<i>Luxilus cardinalis</i>	Cyprinidae	Competitor
Carmine Shiner	<i>Notropis percobromus</i>	Cyprinidae	Competitor
Central Stoneroller	<i>Campostoma anomalum</i>	Cyprinidae	Competitor, Facilitator
Channel Catfish	<i>Ictalurus punctatus</i>	Ictaluridae	Predator, Competitor
Channel Darter	<i>Percina copelandi</i>	Percidae	Competitor
Channel Shiner	<i>Notropis wickliffi</i>	Cyprinidae	Competitor
Common Carp	<i>Cyprinus carpio</i>	Cyprinidae	Competitor
Common Shiner	<i>Luxilus cornutus</i>	Cyprinidae	Predator, Competitor
Creek Chub	<i>Semotilus atromaculatus</i>	Cyprinidae	Predator, Competitor, Facilitator
Emerald Shiner	<i>Notropis atherinoides</i>	Cyprinidae	Competitor
Egg-mimic Darter	<i>Etheostoma pseudocheneis</i>	Percidae	--
Fantail Darter	<i>Etheostoma flabellare</i>	Percidae	Competitor
Fathead Minnow	<i>Pimephales promelas</i>	Cyprinidae	Competitor
Flathead Catfish	<i>Pylodictis olivaris</i>	Ictaluridae	Predator, Competitor
Flathead Chub	<i>Platygobio gracilis</i>	Cyprinidae	Competitor
Freshwater Drum	<i>Aplodinotus grunniens</i>	Sciaenidae	Predator, Competitor
Ghost Shiner	<i>Notropis buechanani</i>	Cyprinidae	Competitor
Gizzard Shad	<i>Dorosoma cepedianum</i>	Clupeidae	Competitor
Golden Redhorse	<i>Moxostoma erythrurum</i>	Catostomidae	Competitor
Golden Shiner	<i>Notemigonus crysoleucas</i>	Cyprinidae	Competitor
Goldeye	<i>Hiodon alosoides</i>	Hiodontidae	Competitor
Goldfish	<i>Carassius auratus</i>	Cyprinidae	Competitor
Grass Carp	<i>Ctenopharyngodon idella</i>	Cyprinidae	Competitor
Gravel Chub	<i>Erimystax x-punctatus</i>	Cyprinidae	Competitor
Green Sunfish	<i>Lepomis cyanellus</i>	Centrarchidae	Predator, Competitor, Facilitator

Appendix B Table 1. List of fish species recorded in the Kansas Stream Assessment and Monitoring dataset used in the calculation of species richness in analyses. Included are common name, scientific name, taxonomic family, and assigned trait guild (if applicable). Species are listed with trait guilds of potential competitors (with invertivore diets), predators (large piscivores), and facilitators (nest building and egg guarding) of Topeka Shiner as assigned by the FishTraits Database and expert advice. Species that are not assigned to a guild are left blank (--).

Common name	Scientific name	Family	Trait Guilds
Greenside Darter	<i>Etheostoma blennioides</i>	Percidae	Competitor
Highfin Carpsucker	<i>Carpionodes velifer</i>	Catostomidae	Competitor
Highland Darter	<i>Etheostoma alta</i>	Percidae	Competitor
Hornyhead Chub	<i>Nocomis biguttatus</i>	Cyprinidae	Predator, Competitor
Inland Silverside	<i>Menidia audens</i>	Atherinopsidae	Competitor
Johnny Darter	<i>Etheostoma nigrum</i>	Percidae	Competitor
Lake Sturgeon	<i>Acipenser fulvescens</i>	Acipenseridae	Competitor
Largemouth Bass	<i>Micropterus nigricans</i>	Centrarchidae	Predator, Competitor
Longear Sunfish	<i>Lepomis megalotis</i>	Centrarchidae	Competitor, Facilitator
Longnose Dace	<i>Rhinichthys cataractae</i>	Cyprinidae	Competitor
Longnose Gar	<i>Lepisosteus osseus</i>	Lepisosteidae	Predator, Competitor
Mooneye	<i>Hiodon tergisus</i>	Hiodontidae	Competitor
Mrigal Carp	<i>Cirrhinus mrigala</i>	Cyprinidae	--
Neosho Madtom	<i>Noturus placidus</i>	Ictaluridae	--
Northern Hogsucker	<i>Hypentelium nigricans</i>	Catostomidae	Competitor
Northern Pike	<i>Esox lucius</i>	Esocidae	Competitor
Northern Plains Killifish	<i>Fundulus kansae</i>	Fundulidae	--
Northern Studfish	<i>Fundulus catenatus</i>	Fundulidae	Competitor
Orangespotted Sunfish	<i>Lepomis humilis</i>	Centrarchidae	Facilitator
Orangethroat Darter	<i>Etheostoma spectabile</i>	Percidae	Competitor
Ozark Logperch	<i>Percina fulvitaenia</i>	Percidae	Competitor
Ozark Minnow	<i>Notropis nubilus</i>	Cyprinidae	Competitor
Paddlefish	<i>Polyodon spathula</i>	Polyodontidae	Competitor
Pallid Sturgeon	<i>Scaphirhynchus albus</i>	Acipenseridae	Competitor
Peppered Chub	<i>Macrhybopsis tetranema</i>	Cyprinidae	--
Plains Killifish	<i>Fundulus zebrinus</i>	Fundulidae	Competitor
Plains Minnow	<i>Hybognathus placitus</i>	Cyprinidae	--
Plains Topminnow	<i>Fundulus sciadicus</i>	Fundulidae	--

Appendix B Table 1. List of fish species recorded in the Kansas Stream Assessment and Monitoring dataset used in the calculation of species richness in analyses. Included are common name, scientific name, taxonomic family, and assigned trait guild (if applicable). Species are listed with trait guilds of potential competitors (with invertivore diets), predators (large piscivores), and facilitators (nest building and egg guarding) of Topeka Shiner as assigned by the FishTraits Database and expert advice. Species that are not assigned to a guild are left blank (--).

Common name	Scientific name	Family	Trait Guilds
Pugnose Minnow	<i>Opsopoeodus emiliae</i>	Cyprinidae	Competitor
Quillback	<i>Carpiodes cyprinus</i>	Catostomidae	Competitor
Rainbow Trout	<i>Oncorhynchus mykiss</i>	Salmonidae	Predator, Competitor
Red River Pupfish	<i>Cyprinodon rubrofluviatilis</i>	Cyprinodontidae	--
Red River Shiner	<i>Notropis bairdi</i>	Cyprinidae	Competitor
Red Shiner	<i>Cyprinella lutrensis</i>	Cyprinidae	Competitor
Redear Sunfish	<i>Lepomis microlophus</i>	Centrarchidae	Competitor
Redfin Shiner	<i>Lythrurus umbratilis</i>	Cyprinidae	Competitor
River Carpsucker	<i>Carpiodes carpio</i>	Catostomidae	Competitor
River Darter	<i>Percina shumardi</i>	Percidae	Competitor
River Redhorse	<i>Moxostoma carinatum</i>	Catostomidae	Competitor
River Shiner	<i>Notropis blennioides</i>	Cyprinidae	Competitor
Rock Bass	<i>Ambloplites rupestris</i>	Centrarchidae	Competitor
Sand Shiner	<i>Notropis stramineus</i>	Cyprinidae	Competitor
Sauger	<i>Sander canadensis</i>	Percidae	Competitor
Shoal Chub	<i>Macrhybopsis hyostoma</i>	Cyprinidae	--
Shorthead Redhorse	<i>Moxostoma macrolepidotum</i>	Catostomidae	Competitor
Shortnose Gar	<i>Lepisosteus platostomus</i>	Lepisosteidae	Predator, Competitor
Shovelnose Sturgeon	<i>Scaphirhynchus platyrhynchus</i>	Acipenseridae	Competitor
Sicklefin Chub	<i>Macrhybopsis meeki</i>	Cyprinidae	Competitor
Silver Carp	<i>Hypophthalmichthys molitrix</i>	Cyprinidae	Competitor
Silver Chub	<i>Macrhybopsis storeriana</i>	Cyprinidae	Competitor
Silverband Shiner	<i>Notropis shumardi</i>	Cyprinidae	Competitor
Skipjack Herring	<i>Alosa chrysochloris</i>	Clupeidae	Competitor
Slender Madtom	<i>Noturus exilis</i>	Ictaluridae	Competitor
Slenderhead Darter	<i>Percina phoxocephala</i>	Percidae	Competitor

Appendix B Table 1. List of fish species recorded in the Kansas Stream Assessment and Monitoring dataset used in the calculation of species richness in analyses. Included are common name, scientific name, taxonomic family, and assigned trait guild (if applicable). Species are listed with trait guilds of potential competitors (with invertivore diets), predators (large piscivores), and facilitators (nest building and egg guarding) of Topeka Shiner as assigned by the FishTraits Database and expert advice. Species that are not assigned to a guild are left blank (--).

Common name	Scientific name	Family	Trait Guilds
Slim Minnow	<i>Pimephales tenellus</i>	Cyprinidae	Competitor
Slough Darter	<i>Etheostoma gracile</i>	Percidae	Competitor
Smallmouth Bass	<i>Micropterus dolomieu</i>	Centrarchidae	Predator, Competitor
Smallmouth Buffalo	<i>Ictiobus bubalus</i>	Catostomidae	Competitor
Southern Redbelly Dace	<i>Chrosomus erythrogaster</i>	Cyprinidae	Competitor
Speckled Chub	<i>Macrhybopsis aestivalis</i>	Cyprinidae	--
Spotfin Shiner	<i>Cyprinella spiloptera</i>	Cyprinidae	Competitor
Spotted Bass	<i>Micropterus punctulatus</i>	Centrarchidae	Predator, Competitor
Spotted Sucker	<i>Minytrema melanops</i>	Catostomidae	Competitor
Stonecat	<i>Noturus flavus</i>	Ictaluridae	Predator, Competitor
Striped Bass	<i>Morone saxatilis</i>	Moronidae	Competitor
Striped Shiner	<i>Luxilus chrysocephalus</i>	Cyprinidae	Competitor
Sturgeon Chub	<i>Macrhybopsis gelida</i>	Cyprinidae	Competitor
Suckermouth Minnow	<i>Phenacobius mirabilis</i>	Cyprinidae	Competitor
Sunburst Darter	<i>Etheostoma eos</i>	Percidae	--
Tadpole Madtom	<i>Noturus gyrinus</i>	Ictaluridae	Competitor
Threadfin Shad	<i>Dorosoma petenense</i>	Clupeidae	Competitor
Walleye	<i>Sander vitreus</i>	Percidae	Predator, Competitor
Warmouth	<i>Lepomis gulosus</i>	Centrarchidae	Predator, Competitor
Western Blacknose Dace	<i>Rhinichthys obtusus</i>	Cyprinidae	--
Western Mosquitofish	<i>Gambusia affinis</i>	Poeciliidae	Competitor
Western Silvery Minnow	<i>Hybognathus argyritis</i>	Cyprinidae	Competitor
White Bass	<i>Morone chrysops</i>	Moronidae	Predator, Competitor
White Crappie	<i>Pomoxis annularis</i>	Centrarchidae	Predator, Competitor
White Perch	<i>Morone americana</i>	Moronidae	Predator, Competitor
White Sucker	<i>Catostomus commersonii</i>	Catostomidae	Competitor
Yellow Bass	<i>Morone mississippiensis</i>	Moronidae	Competitor

Appendix B Table 1. List of fish species recorded in the Kansas Stream Assessment and Monitoring dataset used in the calculation of species richness in analyses. Included are common name, scientific name, taxonomic family, and assigned trait guild (if applicable). Species are listed with trait guilds of potential competitors (with invertivore diets), predators (large piscivores), and facilitators (nest building and egg guarding) of Topeka Shiner as assigned by the FishTraits Database and expert advice. Species that are not assigned to a guild are left blank (--).

Common name	Scientific name	Family	Trait Guilds
Yellow Bullhead	<i>Ameiurus natalis</i>	Ictaluridae	Predator, Competitor
Yellow Perch	<i>Perca flavescens</i>	Percidae	Competitor

Appendix B Table 2. *Objective 1 – Step 3.* Summary of ecological variables investigated in studies of Topeka Shiner in Kansas and other states arranged by year and author names. Included are the study system, type of paper contribution, and ecological variables investigated if relevant. The direction of statistically significant trends in (+/-). Variables shown include *stream size*, *substrate*, *groundwater* and *stream flow*, *biota*, *gradient*, *development agriculture*, and *dams*. Web of Science keywords “Topeka Shiner” (“*Notropis topeka*”) and “Kansas.” The alphabetical labels correspond with Fig. 2. Articles with ecological variables beyond the scope of our study are left blank (--).

	Authors	Year	Location	System	Paper Type	Ecological Variables
<i>a</i>	Mayden & Gilbert	1989	USA	Lab	Taxonomy	--
<i>b</i>	Schmidt & Gold	1995	USA	Lab	Empirical	Biota
<i>c</i>	Adams, Hoover, & Killgore	2000	KS	Lab	Empirical	Flow
<i>d</i>	Hatch & Besaw	2001	MN	River	Empirical	Biota
<i>e</i>	Schrank et al.	2001	KS	Stream	Empirical	Biota (-), Dams, Stream Size (-)
<i>f</i>	Winston	2002	MO	Stream	Empirical	Biota, Gradient, Stream Size (-)
<i>g</i>	Bayless et al.	2003	MO	Stream	Empirical	Gradient, Groundwater, Stream Size
<i>h</i>	Dodds et al.	2004	USA	Stream	Synthesis	--
<i>i</i>	Wall & Berry	2004	USA	Stream	Review	Gradient, Groundwater, Stream Size, Substrate
<i>j</i>	Wall et al.	2004	SD	Stream	Model	Agriculture (+), Flow (+), Gradient (-), Groundwater (+), Stream Size (+), Substrate (+)
<i>k</i>	Knight & Gido	2005	KS	Lab	Empirical	Biota
<i>l</i>	Oakes et al.	2005	KS & NE	Stream	Model	--
<i>m</i>	Koehle & Adelman	2007	MO & KS	Lab	Empirical	--
<i>n</i>	Anderson & Sarver	2008	SD	Lab	Empirical	--
<i>o</i>	Adelman et al.	2009	MO & MN	Lab	Empirical	--
<i>p</i>	Pullen et al.	2009	KS	Stream	Empirical	Biota
<i>q</i>	Witte et al.	2009	MO & KS	Lab	Empirical	Substrate (+)
<i>r</i>	Bouska & Paukert	2010	KS	Stream	Empirical	Flow

Appendix B Table 2. *Objective 1 – Step 3.* Summary of ecological variables investigated in studies of Topeka Shiner in Kansas and other states arranged by year and author names. Included are the study system, type of paper contribution, and ecological variables investigated if relevant. The direction of statistically significant trends in (+/-). Variables shown include *stream size*, *substrate*, *groundwater* and *stream flow*, *biota*, *gradient*, *development agriculture*, and *dams*. Web of Science keywords “Topeka Shiner” (“*Notropis topeka*”) and “Kansas.” The alphabetical labels correspond with Fig. 2. Articles with ecological variables beyond the scope of our study are left blank (--).

	Authors	Year	Location	System	Paper Type	Ecological Variables
<i>s</i>	Bakevich, Pierce, & Quist	2013	IA	Stream & Off Channel	Empirical	Biota
<i>t</i>	Gerken & Paukert	2013	KS	Stream	Empirical	Biota (+), Dams, Development, Stream Size, Substrate
<i>u</i>	Bakevich, Pierce, & Quist	2015	IA	Stream & Off Channel	Review	--
<i>v</i>	Fischer et al.	2018	IA	Oxbow	Empirical	Stream Size
<i>w</i>	Bartell et al.	2019	IA, NE, MO	Stream	Model	Agriculture
<i>x</i>	Pierce et al.	2019	IA	Watersheds	Qualitative	--
<i>y</i>	Simpson et al.	2019	IA & MN	Oxbow	Empirical	Biota (+), Stream Size (+)
<i>z</i>	Zambory et al.	2019	IA & MN	Oxbow	Model	--
<i>aa</i>	Mott et al.	2021	MO	Lab	Empirical	--
<i>ab</i>	Schmolke et al.	2021	IA, NE, MO	Oxbow	Model	Agriculture, Biota
<i>ac</i>	Simpson et al.	2021	IA	Stream	Empirical	Biota (-/+)
<i>ad</i>	Osterhaus et al.	2022	IA & MN	Oxbow	Model	Biota
<i>ae</i>	Galic et al.	2023	IA, NE, MO	Oxbow	Model	Stream Size (+)
<i>af</i>	Pfaff, Hase, & Gido	2023	KS	Pond	Empirical	Biota
<i>ag</i>	Coulter et al.	2024	USA	Stream	Synthesis	Biota
<i>ah</i>	Wahl et al.	2024	IA & MN	Stream	Model	Stream Size (-)

Appendix B Table 3. Objective 1 – Step 8. Results from multiple logistic regression on original data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas (n=670). Eighty-one percent of the data was collected between 2000-2021. Included are variable name, coefficient (b), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor ecological variables include: (A) *stream order* (modified Strahler), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). These calculations correspond to Step 8F of the Rode et al. 2025 framework. $\Delta AIC_c = 3.22$. The new variable metric added is underlined.

Variable	B	SE	Z	95% CI		p		OR	VIF
				Lower	Upper				
(Intercept)	-6.40	3.83	-1.67	-13.92	1.11	0.09		0.00	
<i>Clay</i>	0.18	0.02	8.55	0.14	0.22	0.00	*	1.20	1.29
<u><i>Stream Order</i></u>	<u>-0.44</u>	<u>0.13</u>	<u>-3.33</u>	<u>-0.70</u>	<u>-0.18</u>	<u>0.00</u>	<u>*</u>	<u>0.64</u>	<u>2.08</u>
<i>Number of Dams</i>	-0.10	0.02	-4.73	-0.14	-0.06	0.00	*	0.90	1.78
<i>Species Richness</i>	-0.09	0.02	-4.89	-0.13	-0.05	0.00	*	0.91	1.26
<i>Development</i>	-5.29	2.92	-1.81	-11.01	0.42	0.07		0.01	1.41
<i>Grassland</i>	1.66	1.20	1.38	-0.70	4.01	0.17		5.24	10.96
<i>Pastureland</i>	-2.15	1.99	-1.08	-6.04	1.74	0.28		0.12	2.90
<i>Baseflow Index</i>	-0.03	0.03	-1.02	-0.08	0.03	0.31		0.97	1.28
<i>Curvature</i>	0.12	0.13	0.96	-0.13	0.37	0.34		1.13	1.20
<i>Cropland</i>	-0.08	1.29	-0.06	-2.60	2.44	0.95		0.92	6.96
<i>Gradient</i>	-2.92	38.33	-0.08	-78.04	72.20	0.94		0.05	2.24

Appendix B Table 4. *Objective 1 – Step 8.* Results from multiple logistic regression on original data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas (n=670). Eighty-one percent of the data was collected between 2000-2021. Included are variable name, coefficient (b), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor ecological variables include: (A) *stream velocity* (average annual gage estimate in feet per second), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). These calculations correspond to Step 8F of the Rode et al. 2025 framework. $\Delta AIC_c = 6.24$. The new variable metric added is underlined.

Variable	β	SE	Z	95% CI		p		OR	VIF
				Lower	Upper				
(Intercept)	-5.70	4.04	-1.41	-13.61	2.22	0.16		0.00	
<i>Clay</i>	0.18	0.02	8.36	0.14	0.22	0.00	*	1.20	1.29
<u><i>Stream Velocity</i></u>	<u>-1.66</u>	<u>0.65</u>	<u>-2.53</u>	<u>-2.94</u>	<u>-0.37</u>	<u>0.01</u>	<u>*</u>	<u>0.19</u>	<u>1.49</u>
<i>Number of Dams</i>	-0.11	0.02	-5.14	-0.15	-0.07	0.00	*	0.90	1.77
<i>Species Richness</i>	-0.09	0.02	-4.85	-0.13	-0.05	0.00	*	0.91	1.27
<i>Development</i>	-4.49	2.89	-1.55	-10.15	1.18	0.12		0.01	1.36
<i>Grassland</i>	2.24	1.16	1.93	-0.03	4.52	0.05		9.44	10.29
<i>Pastureland</i>	-1.92	2.01	-0.96	-5.86	2.01	0.34		0.15	2.84
<i>Baseflow Index</i>	-0.04	0.03	-1.32	-0.10	0.02	0.19		0.96	1.27
<i>Curvature</i>	0.10	0.13	0.77	-0.15	0.35	0.44		1.10	1.20
<i>Cropland</i>	0.62	1.24	0.50	-1.81	3.05	0.62		1.85	6.63
<i>Gradient</i>	39.12	35.14	1.11	-29.76	107.99	0.27		9.73E+16	1.78

Appendix B Table 5. Objective 2B – Step 8. Results from multiple logistic regression on original data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas (n=670). Eighty-one percent of the data was collected between 2000-2021. Included are variable name, coefficient (b), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor ecological variables include: (A) *stream depth* (meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). These calculations correspond to Step 8F of the Rode et al. 2025 framework. The new variable metric added is underlined.

Variable	β	SE	Z	95% CI		p		OR	VIF
				Lower	Upper				
(Intercept)	-23.08	6.50	-3.55	-35.82	-10.34	0.00	*	0.00	
<i>Clay</i>	0.12	0.02	5.21	0.07	0.16	0.00	*	1.13	1.28
<i>Stream Depth</i>	-0.71	0.26	-2.75	-1.22	-0.20	0.01	*	0.49	1.76
<i>Number of Dams</i>	-0.12	0.02	-5.08	-0.16	-0.07	0.00	*	0.89	1.95
<i>Species Richness</i>	-0.09	0.02	-4.44	-0.13	-0.05	0.00	*	0.92	1.35
<i>Development</i>	-5.45	3.00	-1.82	-11.32	0.43	0.07		0.00	1.38
<i>Grassland</i>	1.95	1.13	1.72	-0.27	4.17	0.08		7.02	8.94
<i>Pastureland</i>	-0.43	1.97	-0.22	-4.29	3.42	0.83		0.65	2.72
<i>Baseflow Index</i>	1.65	0.46	3.61	0.75	2.55	0.00	*	5.21	113.67
<u><i>Baseflow Index</i>²</u>	<u>-0.03</u>	<u>0.01</u>	<u>-3.53</u>	<u>-0.05</u>	<u>-0.02</u>	<u>0.00</u>	<u>*</u>	<u>0.97</u>	<u>113.58</u>
<i>Curvature</i>	0.07	0.14	0.53	-0.19	0.34	0.59		1.07	1.22
<i>Cropland</i>	0.24	1.23	0.20	-2.17	2.65	0.85		1.27	5.91
<i>Gradient</i>	-25.71	38.75	-0.66	-101.66	50.24	0.51		0.00	2.15

Appendix B Table 6. *Objective 2C – Step 8.* Results from multiple logistic regression on original data using 11 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas (n=670). Eighty-one percent of the data was collected between 2000-2021. Included are variable name, coefficient (b), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor ecological variables include: (A) *stream depth* (meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *predator species richness* (number of fish species in a single sample which are potential predators of Topeka Shiner), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). These calculations correspond to Step 8F of the Rode et al. 2025 framework. The new variable metric added is underlined.

Variable	β	SE	Z	95% CI		p		OR	VIF
				Lower	Upper				
(Intercept)	-22.68	6.45	-3.52	-35.33	-10.04	0.00	*	0.00	
<i>Clay</i>	0.12	0.02	5.05	0.07	0.16	0.00	*	1.12	1.26
<i>Stream Depth</i>	-0.77	0.27	-2.82	-1.30	-0.23	0.00	*	0.46	1.79
<i>Number of Dams</i>	-0.13	0.02	-5.44	-0.18	-0.08	0.00	*	0.88	2.03
<u><i>Predator Species Richness</i></u>	<u>-0.32</u>	<u>0.06</u>	<u>-4.90</u>	<u>-0.44</u>	<u>-0.19</u>	<u>0.00</u>	<u>*</u>	<u>0.73</u>	<u>1.39</u>
<i>Development</i>	-5.20	3.10	-1.68	-11.27	0.88	0.09		0.01	1.41
<i>Grassland</i>	2.34	1.18	1.99	0.03	4.65	0.05	*	10.37	9.38
<i>Pastureland</i>	-0.19	2.02	-0.10	-4.16	3.77	0.92		0.82	2.78
<i>Baseflow Index</i>	1.62	0.45	3.59	0.73	2.50	0.00	*	5.03	110.25
<i>Baseflow Index</i> ²	-0.03	0.01	-3.52	-0.05	-0.01	0.00	*	0.97	110.51
<i>Curvature</i>	0.08	0.14	0.58	-0.19	0.34	0.56		1.08	1.22
<i>Cropland</i>	0.61	1.28	0.48	-1.89	3.12	0.63		1.84	6.24
<i>Gradient</i>	-26.48	38.97	-0.68	-102.85	49.90	0.50		0.00	2.15

Appendix B Table 7. *Objective 2D – Step 8.* Results from multiple logistic regression on original data using 5 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas (n=670). Eighty-one percent of the data was collected between 2000-2021. Included are variable name, coefficient (b), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF). Predictor ecological variables include: (A) *stream depth* (meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), and (E) *number of dams* (total in the mainstem). These calculations correspond to Step 8F of the Rode et al. 2025 framework. The new interaction between *clay* and *baseflow index* added is underlined.

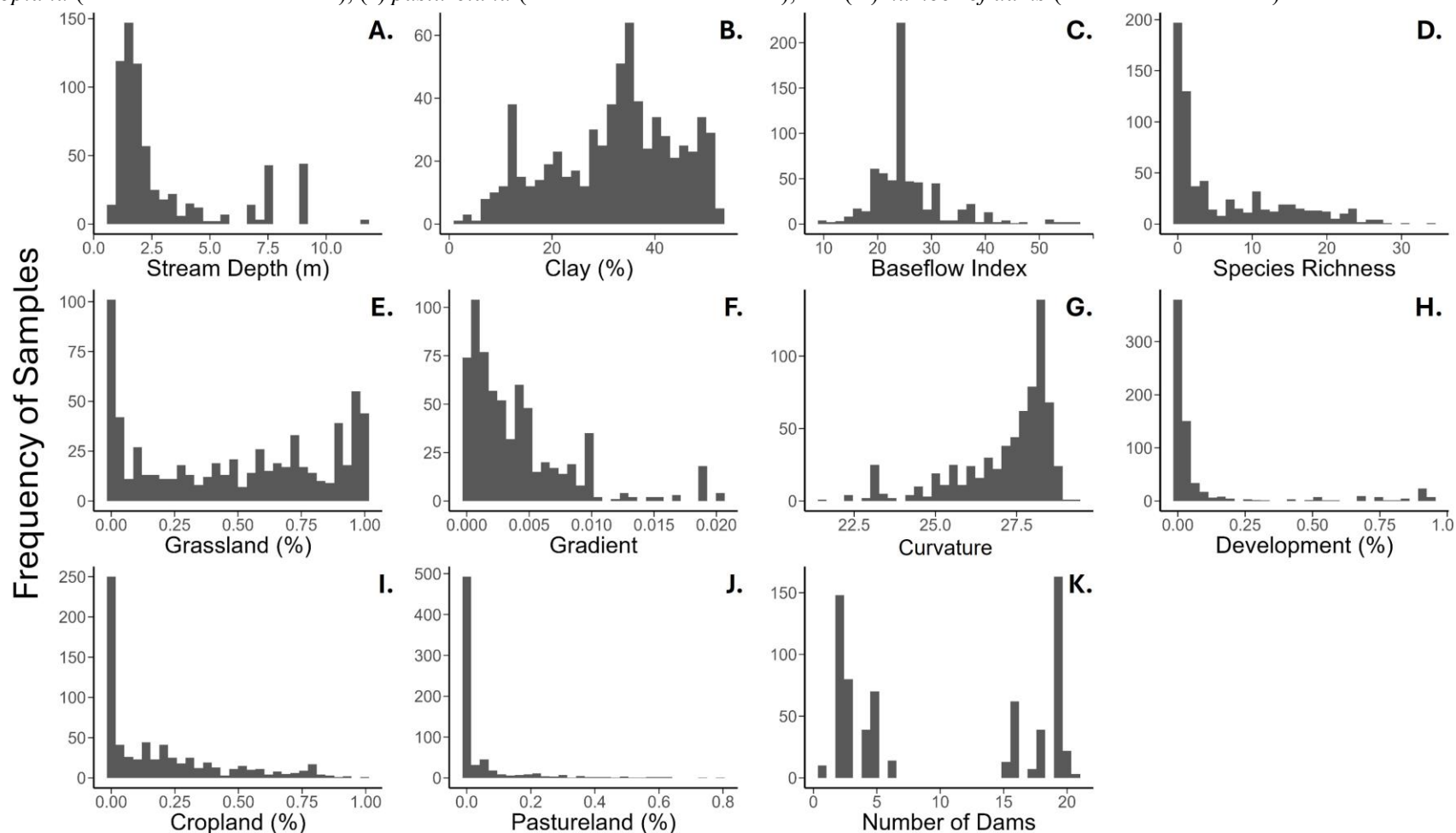
Variable	β	SE	Z	95% CI		p		OR	VIF
				Lower	Upper				
(Intercept)	-34.62	9.66	-3.58	-53.55	-15.69	0.00	*	0.00	
<i>Clay</i>	0.45	0.17	2.69	0.12	0.78	0.01	*	1.57	74.93
<i>Stream Depth</i>	-1.04	0.25	-4.09	-1.54	-0.54	0.00	*	0.35	1.29
<i>Number of Dams</i>	-0.08	0.02	-4.37	-0.12	-0.05	0.00	*	0.92	1.48
<i>Predator Species Richness</i>	-0.28	0.06	-4.74	-0.40	-0.16	0.00	*	0.76	1.27
<i>Baseflow Index</i>	2.34	0.63	3.73	1.11	3.56	0.00	*	10.36	252.72
<i>Baseflow Index</i> ²	-0.04	0.01	-3.87	-0.06	-0.02	0.00	*	0.96	149.53
<u><i>Clay: Baseflow Index</i></u>	<u>-0.01</u>	<u>0.01</u>	<u>-1.96</u>	<u>-0.03</u>	<u>0.00</u>	<u>0.05</u>	<u>*</u>	<u>0.99</u>	<u>96.47</u>

Appendix B Table 8. Objective 2E – Step 8. Results from multiple logistic regression on original data using 5 ecological explanatory variables to model the response of Topeka Shiner probability of presence in Kansas (n=670). Eighty-one percent of the data was collected between 2000-2021. Included are variable name, coefficient (b), standard error (SE), Z-score (Z), lower and upper 95% confidence interval, p-value (asterisk to denote $p \leq 0.05$), odds ratio (OR), and variable inflation factor (VIF) for fixed effects and standard deviation (SD) and 95% CI for random intercepts. Predictor ecological variables include: (A) *stream depth* (meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), and (E) *number of dams* (total in the mainstem). These calculations correspond to Step 8F of the Rode et al. 2025 framework.

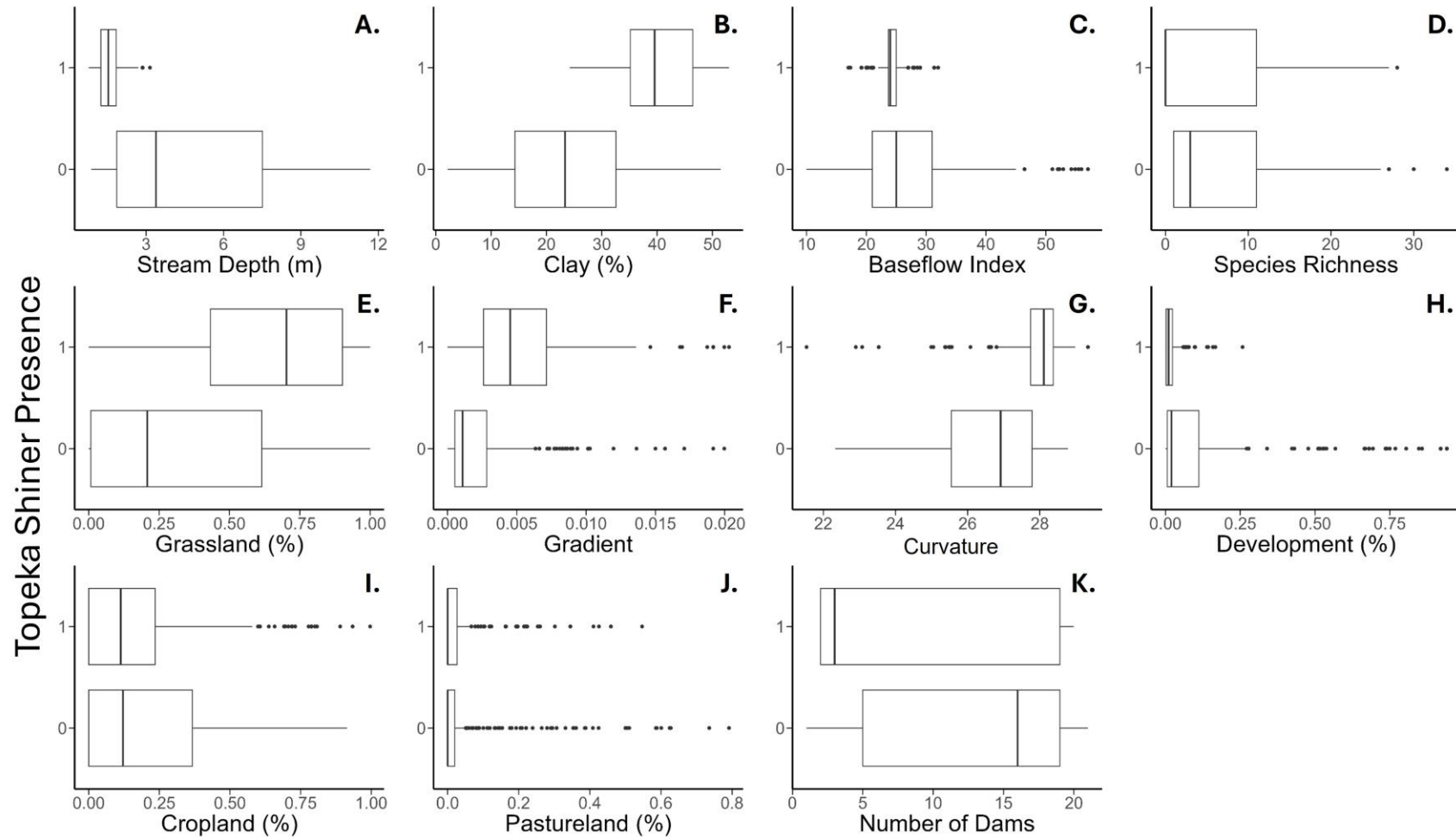
Fixed Effects									
Variable	β	SE	Z	95% CI		p	*	OR	VIF
				Lower	Upper				
(Intercept)	-47.93	14.13	-3.39	-75.62	-20.23	0.00	*	0.00	
<i>Clay</i>	0.63	0.23	2.78	0.19	1.07	0.01	*	1.87	99.95
<i>Stream Depth</i>	-1.10	0.24	-4.55	-1.58	-0.63	0.00	*	0.33	1.31
<i>Number of Dams</i>	0.10	0.18	0.58	-0.24	0.45	0.56		1.11	1.14
<i>Predator Species Richness</i>	-0.30	0.07	-4.49	-0.42	-0.17	0.00	*	0.74	1.05
<i>Baseflow Index</i>	2.45	0.78	3.14	0.92	3.98	0.00	*	11.61	203.84
<i>Baseflow Index</i> ²	-0.03	0.01	-2.94	-0.06	-0.01	0.00	*	0.97	117.25
<i>Clay: Baseflow Index</i>	-0.02	0.01	-2.42	-0.04	0.00	0.02	*	0.98	134.11
Random Effects									
SD									
Basin	7.21			1.79	29.04				

Figures

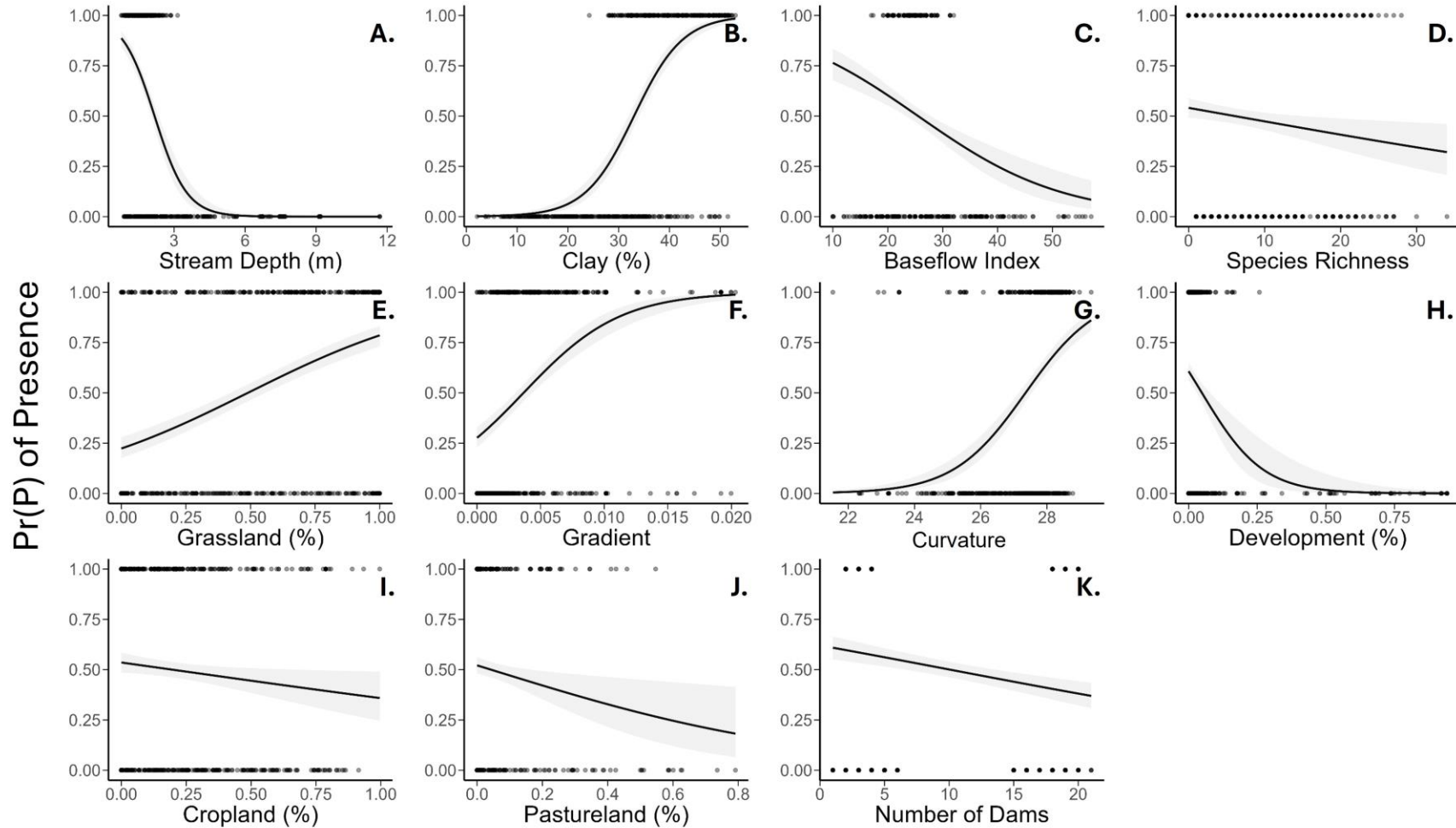
Appendix B Figure 1. *Objective 1 – Step 8B*. Histogram sample frequencies (Y axis) of eleven ecological variables associated with 670 stream site samples in Kansas from 1947 to 2020, including (A) *stream depth* (estimated in meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem).



Appendix B Figure 2. Objective 1 – Step 8D. Boxplots of eleven ecological variables (X axis) including (A) *stream depth* (estimated in meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). Data distributions are split by factor of Topeka Shiner presence (1) and absence (0) on the Y axis from collections in Kansas between 1947 and 2020.



Appendix B Figure 3. Objective 1 – Step 8D. Single logistic regression response trends for the impact of eleven ecological variables including (A) *stream depth* (estimated in meters), (B) *clay* (% in catchment surface geology), (C) *baseflow index* (% contribution of groundwater and other sustained water sources to stream flow), (D) *species richness* (number of fish species in a single sample), (E) *grassland* (% cover in catchment), (F) *gradient* (m/m), (G) *curvature*, (H) *development* (% cover in the catchment), (I) *cropland* (% cover in the catchment), (J) *pastureland* (% cover in the catchment), and (K) *number of dams* (total in the mainstem). The response is the predicted probability of Topeka Shiner presence (Y axis) in Kansas from 1947 to 2020.



Appendix B Figure 4. Objective 2E – Step 8F. Intercept adjustments (X axis) for the random effect of twelve major river basins (Y axis) in KS to the modeled predicted probability of Topeka Shiner presence from 1947 to 2020.

