Summer 8-18-2023

Integrating Environmental DNA, Traditional Fisheries Techniques, and Species Distribution Modeling to Assess Bridle Shiner Status in Maine

Lara S. Katz
University of Maine, lara.katz@maine.edu

INTEGRATING ENVIRONMENTAL DNA, TRADITIONAL FISHERIES TECHNIQUES, AND SPECIES DISTRIBUTION MODELING TO ASSESS BRIDLE SHINER STATUS IN MAINE

By

Lara S. Katz

B.S. University of Maine, 2015

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

(in Wildlife Ecology)

The Graduate School

The University of Maine

August 2023

Advisory Committee:

Dr. Joseph Zydlewski, Unit Leader, USGS Maine Cooperative Fish & Wildlife Research Unit and Professor, University of Maine, Co-Advisor

Dr. Stephen M. Coghlan, Jr., Associate Professor of Freshwater Fisheries Ecology, University of Maine, Co-Advisor

Dr. Michael Kinnison, Professor of Evolutionary Applications, University of Maine

INTEGRATING ENVIRONMENTAL DNA, TRADITIONAL FISHERIES TECHNIQUES, AND SPECIES DISTRIBUTION MODELING TO ASSESS BRIDLE SHINER STATUS IN MAINE

By Lara S. Katz

Thesis Advisors: Dr. Joseph Zydlewski and Dr. Stephen Coghlan, Jr.

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Master of Science
(in Wildlife Ecology)
August 2023

The bridle shiner (*Notropis bifrenatus*) is a small minnow species native to the eastern United States and southeastern Canada. The species is declining dramatically throughout most of its native range and has legal protection or concern status in thirteen states and two Canadian provinces. In Maine, the bridle shiner is listed as a Species of Special Concern and considered a Species of Greatest Conservation Need, partially because we lack a basic understanding of their status and distribution within the state. Bridle shiners have historically been found in southern and western Maine in densely vegetated, shallow habitats along the shorelines of streams and ponds. Surveys performed at sites where the shiners were once abundant have yielded very few or none of these fish. This project informed the Maine Department of Inland Fisheries & Wildlife on the status of the species in Maine and provides a foundation for future long-term monitoring of bridle shiner populations in the State.

We used a combination of both direct capture techniques and environmental DNA (eDNA) to locate bridle shiners. eDNA is increasingly being used to detect rare aquatic species such as bridle shiners because it is both highly sensitive and less invasive than direct capture. We

designed a single-species primer-probe assay to detect bridle shiner DNA, then surveyed 32 sites with a record of historic bridle shiner occurrence. In addition to collecting eDNA samples (2021-2022), we surveyed 29 sites using traditional seine netting techniques in 2021. In 2022, we used a preliminary habitat suitability model to select 46 locations with unknown bridle shiner presence to survey with eDNA. To refine eDNA methodology, we assessed trends in eDNA detection probability across seasons and compared DNA detection between three filter pore sizes. We rediscovered bridle shiner populations at 11 of 32 historically occupied sites and documented bridle shiners in four additional waterbodies. We determined that eDNA surveys were most effective in early or midsummer, and that larger filter pore sizes are a viable option for surveying bridle shiners.

Species distribution modeling (SDM) statistically associates species occurrence data with environmental variables to evaluate habitat suitability. We used an ensemble species distribution modeling (SDM) approach to identify both the current and historic range of the bridle shiner within Maine and New Hampshire. We also investigated how local habitat characteristics influenced bridle shiner presence using generalized linear models. Both historic site surveys and ensemble SDMs suggest that there has been a substantial loss of historic bridle shiner habitat in Maine (-62%) and New Hampshire (-46%). At the landscape scale, we found significant effects of forest type, catchment position, soil composition, elevation, and slope on bridle shiners. Within a site, bridle shiners were associated with areas that had a higher proportion of complex-leaved submerged aquatic vegetation and a lower proportion of persistent emergent and floating vegetation. We determined that both eDNA and seine net surveys are viable options for monitoring bridle shiners in Maine, and that such survey strategies can be used with species

distribution models to focus future surveys and to identify areas of possible conservation, reintroduction, or restoration actions.

ACKNOWLEDGEMENTS

I am grateful for the funding support of the Maine Department of Inland Fisheries & Wildlife and the Maine Outdoor Heritage Fund. In-kind support was provided by the U.S. Geological Survey Maine Cooperative Fish and Wildlife Research Unit. Several travel grants were provided by the University of Maine Department of Wildlife, Fisheries, and Conservation Biology, the University of Maine College of Natural Sciences, Forestry, and Agriculture, the American Fisheries Society, and the Atlantic International Chapter of the American Fisheries Society. The University of Maine Coordinated Operating Research Entities (CORE) provided laboratory space, equipment, and technicians to process eDNA samples.

There are many people who have supported me over the course of this project, and I am incredibly appreciative of all of them. First and foremost, I would like to thank my advisory committee: Dr. Joseph Zydlewski, Dr. Stephen Coghlan, and Dr. Michael Kinnison for all of their guidance and support over the course of this project. I also sincerely thank Merry Gallagher (Maine Department Inland Fisheries & Wildlife), who was instrumental in getting this project started, funded, and planned, and who provided me with Maine's historic bridle shiner survey data.

I would also like to thank my manuscript co-authors Dr. Erik Blomberg, Geneva York, and Matt Carpenter. Geneva designed the eDNA primer, ran all of my eDNA samples through qPCR, and answered my many questions about eDNA. Erik helped me tremendously with the occupancy modeling and helped me become much more familiar with coding in R. My species distribution modeling chapter would not have been possible without Matt Carpenter (New Hampshire Fish & Game Department) and Dr. Shawn Snyder. Matt provided me with all of New Hampshire's bridle shiner survey data, which made the models much more accurate than they

would have otherwise been. Shawn shared his species distribution modeling expertise and advice when I had zero experience with it, saving me weeks of extra work and confusion.

Dr. Jeremy Wright and Bryan Weatherwax (New York State Museum) made primer development possible by providing us with bridle shiner tissue samples from New York. Merry Gallagher was also critical to this process, as she provided us with shiner and darter tissue samples from other species to help Geneva test the primer.

Many people helped me with the logistics of this project. James Pellerin (MDIFW) and Matt Lubejko (MDIFW) were instrumental in project planning and in finding historic Maine bridle shiner records. Dr. Kasey Pregler and Matt Carpenter gave me advice on where and how to look for bridle shiners. Dr. Tora Johnson and Amy Dowley helped me build my habitat suitability index model to choose sampling sites. Dr. Erin Grey, Dr. Andy Rominger, and Dr. Sue Ishaq introduced me to eDNA methodology and bioinformatics. Rena Carey, Katherine Goodine, and Molly-Jean Langlais-Parker patiently helped me with all the administrative aspects of the project. I would also like to thank the many landowners who allowed me to use their property for lake and pond access and/or who let me use their docks.

I'd like to thank the Zydlewski lab (Ernie Atkinson, Cody Dillingham, Guillermo Figueroa-Muñoz, Melissa Flye, Emilie Hickox, Matt Mensinger, Carolyn Merriam, Erin Peterson, Sarah Rubenstein, Rylee Smith, Sarah Vogel, and Kory Whittum), past and present, for the many ways that they have supported me over the past couple of years. I would especially like to thank Rylee Smith and Cody Dillingham, who have been wonderful friends/officemates and have helped me with many aspects of my coursework and my research.

This project would not have been possible without the help of my two technicians, Emile Gauvin and Maddie Huerth (Wabanaki Youth in Science), and the many students, volunteers,

and friends who helped me in the field. Andrea Casey, Cody Dillingham, Molly Donlan (NPS), Henry Guy, Jakob Hallett, Silvia Hartt, Emilie Hickox, Morgan Ingalls (NPS), Justice Maddocks-Wilbur, Matt Mensinger, Carolyn Merriam, Allie Ouimet, Rylee Smith, and Sarah Vogel assisted with eDNA sample collection and seine netting and tolerated the long drive to and from the sites. Kory Whittum helped me identify young-of-year minnows, and Jeb Young helped make eDNA sampling kits. Another thank-you goes out to the eDNA CORE lab technicians who extracted DNA from my samples and prepared it for qPCR.

I am also grateful to Becky Cole-Will, Bruce Connery, Molly Donlan, Morgan Ingalls, and Bik Wheeler at Acadia National Park for helping me apply to grad school and encouraging me to pursue a masters in the first place. Morgan and Molly also helped me with seine netting in Acadia and win the unofficial award for most fish caught in one day (nearly 1000!).

A special thanks to my parents, sister, and grandmother, who have been an endless source of support for me. I would not have been able to complete this degree without their support and the support of my advisor, Joe, and my friends, who first helped me navigate a difficult loss, then provided patience and support as I recovered from long-COVID.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

AN INTEGRATIVE APPROACH TO ASSESSING BRIDLE SHINERS (*NOTROPIS BIFRENATUS*) USING ENVIRONMENTAL DNA AND TRADITIONAL TECHNIQUES

1.1. Introduction

Freshwater ecosystems are among the most critically threatened habitats on Earth, and within these ecosystems, freshwater fish are among the most imperiled animals (Dudgeon et al. 2006; Jelks et al. 2008). As of 2008, nearly 40% of North America's freshwater and diadromous fish species were considered vulnerable, threatened, or endangered (Jelks et al. 2008). Maine has a relatively low biodiversity of freshwater fish, with only 65 species in total (Everhart 2002; Wick 2007; Gallagher 2010a, 2010b). Nineteen of these species are introduced or exotic, and 46 are native to at least a portion of the State (United States Geological Survey [USGS] 2021). Nearly one quarter of Maine's native freshwater fish species are considered species of conservation concern (Maine Dept. of Inland Fisheries and Wildlife [MDIFW] 2021).

Anthropogenic environmental impacts such as habitat loss, degradation, and fragmentation have increased in prevalence over recent decades (Paul and Meyer 2001; Walsh et al. 2005). Small-bodied fish species such as minnows and darters are especially vulnerable to the impacts of habitat alteration and destruction (Whittier et al. 1997; Olden et al. 2007). Changes to natural hydrologic regimes, increased turbidity and pollution, and the introduction of invasive species frequently pose threats to small-bodied freshwater fishes (Angermeier 1995; Bunn and Arthington 2002; Jelks et al. 2008; Gray et al. 2016). Recent declines and extirpations of native minnows in the northeastern United States and Canada have been linked to increased anthropogenic activity in a watershed and to the stocking of non-native predatory sport fishes

(Whittier et al. 1997). For example, Whittier et al. (Whittier et al. 1997) found that the introduction of predators such as largemouth bass (*Micropterus salmoides*), smallmouth bass (*M. dolomieu*), and northern pike (*Esox lucius*) was the most consistent factor related to declines in native minnow species richness in northeastern lakes.

Small-bodied fish are especially threatened with extirpation when they occupy limited geographic ranges and narrow ecological niches. These specialist species likely already have lower populations to begin with and are thus especially vulnerable to disturbance and environmental stochasticity (Angermeier 1995). Shifts in water chemistry, temperature, and vegetation resulting from anthropogenic impacts can all result in local extirpations of specialist species (Angermeier 1995). One such small-bodied, specialist minnow is the bridle shiner (*Notropis bifrenatus*). Native to the eastern United States and Canada, bridle shiners inhabit clear, slow-moving waters in lakes, ponds, streams, and smaller rivers (Page and Burr 2011). They are commonly found in wetland habitats that support beds of submerged aquatic plants (Harrington 1948a; Jensen and Vokoun 2013). Bridle shiners grow to a maximum of approximately 60-mm in length and can live to be about two years old (Harrington 1948a; Committee on the Status of Endangered Wildlife in Canada [COSEWIC] 2013). Bridle shiners spawn between May and August when water temperatures are between 14 and 27°C (COSEWIC 2013).

Limited data suggest that the bridle shiner has been declining dramatically throughout most of its native range (Pregler et al. 2015), and probably has been extirpated entirely from the District of Columbia (Hammerson 2021) and Maryland (Kilian et al. 2011). This species now receives concern status or legal protection in thirteen states and two provinces (COSEWIC 2013;

Hammerson 2021). Bridle shiners are listed as a Species of Special Concern and considered a Species of Greatest Conservation Need in Maine (MDIFW 2015, 2021).

Bridle shiner declines are likely due to the same factors that affect other minnow species, especially habitat loss and degradation. These fish are vulnerable to disturbances such as lake drawdowns and herbicide use because they live on the shoreline and require access to abundant vegetation (Pregler et al. 2019). Bridle shiners historically occurred in regions of Maine where freshwater systems are more heavily degraded and stressed by human population growth, habitat loss, and climate change. Bridle shiner declines are therefore suspected in the State, but we lack the basic ecological data on their distribution and abundance necessary to assess their status.

Low detection probabilities are a concern when monitoring this and other rare, small-bodied species. Traditional fisheries techniques such as backpack electrofishing tend to have low capture probabilities for rare organisms and are therefore more useful when trying to detect abundant species (Jerde et al. 2011). Because of their distinct habitat requirements, effective bridle shiner sampling presents unique logistical constraints. Seine nets have successfully been used to capture bridle shiners and other minnows in clear, slow-moving water with abundant vegetation (Jensen and Vokoun 2013; Pregler et al. 2015; Lamothe and Drake 2020), but seine netting is labor intensive.

Environmental DNA (eDNA) offers an alternative technique. eDNA methods have been successfully applied in the study of other rare fish (e.g., Jerde et al. 2011; Hinlo et al. 2018; Robinson et al. 2019), and are both highly sensitive to the DNA of the target organism (Turner et al. 2014) and non-invasive (Valentini et al. 2016; Deiner et al. 2016). Once shed, DNA breaks down in the environment and has a limited period of availability for detection. Aquatic eDNA degrades relatively quickly, and provides a near-current snapshot of species presence (Dejean et

al. 2011; Thomsen et al. 2012; Agersnap et al. 2022). Therefore, the utility of eDNA depends on the sensitivity and specificity of the extraction method, but also on the sampling approach with regards to the location, timing, and extent of sampling. One way to increase eDNA detection probability is to collect samples during times of the year when eDNA quantities will be higher, such as during the spawning season (de Souza et al. 2016).

There are tradeoffs associated with each season when sampling eDNA. During summer, bridle shiner habitat is more easily identifiable, and fish shed more DNA (Lacoursière-Roussel et al. 2016b), but the collected water contains algae, bacteria, and plant material that is difficult to filter. eDNA also degrades faster at higher water temperatures (Barnes et al. 2014; Eichmiller et al. 2016; Goldberg et al. 2018). In winter, water samples filter quickly, but detection probability of DNA might be lower due to seasonal changes in fish behavior (de Souza et al. 2016; Thalinger et al. 2021), metabolism (Lacoursière-Roussel et al. 2016b), and stream discharge (Thalinger et al. 2021).

The goal of this study was to characterize the population status and distribution of the bridle shiner in Maine and provide a method for long-term assessment. Focal areas included watersheds in southern and western Maine that represent the northeastern extent of the species' range in the US. The objectives of the study were to 1) survey areas that have previously supported bridle shiner populations in Maine using eDNA in conjunction with traditional seine netting methods, 2) use eDNA sampling to survey areas with unknown bridle shiner presence, and 3) refine eDNA methodology for future surveys.

1.2. Methods

1.2.1. Study Area

Bridle shiners were historically found in southern and western Maine in densely vegetated, shallow habitats along the shorelines of streams and ponds (Cooper 1939). The earliest MDIFW records of bridle shiners in the State were documented by Kendall (1914) and Cooper (1939), and subsequent records were the result of incidental captures during various stream and lake surveys (Doering et al. 1995; Yoder et al. 2009; Gallagher 2010a, 2010b; U.S. Environmental Protection Agency [USEPA] 2016).

In total, bridle shiners were reported at 38 locations between 1937 and 2010 (Figure 1.1). All the lake and pond sites (n = 16) have been resurveyed at least once between 1939 and 2021, but bridle shiners have only been captured in two ponds since 1992 (USEPA 2016). Bridle shiners appear on the species list of four lakes in central and eastern Maine (open circles; Figure 1.1), but MDIFW biologists consider these records to be misidentifications (most likely of the closely related blacknose shiner, *N. heterolepis*). Doering et al. (1995) found one bridle shiner during a survey of Marshall Brook at Acadia National Park and Stone et al. (2001) hypothesized that this may have been a bait-bucket introduction (Figure 1.1).

Although bridle shiners are considered a lake-dwelling species (Whittier et al. 1997), almost all the recent (1990-2010) bridle shiner documentations in Maine have come from streams and rivers (n = 8 records). Road crossing coordinates were available for the 22 stream sites with records of bridle shiner presence, but historic survey coordinates were not available for the 12 lakes and ponds with credible records (Kendall 1914; Cooper 1939; Doering et al. 1995; Yoder et al. 2010; Gallagher 2010a, 2010b; USEPA 2016). Surveys performed at sites where the

shiners were once abundant have yielded very few or none of these fish, but surveys specifically targeting bridle shiners had not been conducted prior to this study.

While bridle shiners have not been captured at any of their historically occupied sites in recent years, this may reflect biases in gear, sampled habitat, and survey goals rather than true absences. For example, MDIFW lake and pond surveys between 1950 and 1970 often did not distinguish between minnow species, and sampling equipment used sizes that allowed smaller fish to evade capture (Stone et al. 2001). Therefore, it is possible that bridle shiner populations remain at some or all of these historic sites, but the overall status of bridle shiner populations in Maine is unknown.

We surveyed bridle shiners at 95 locations within 68 waterbodies in southwestern Maine, USA between June 2021 and November 2022. We visited an additional three sites where we were unable to sample with eDNA or a seine net (i.e., could not get permission to sample, $n = 2$; pond dried and could no longer support fish, $n = 1$). Sampling in both years focused on the Saco, Presumpscot, and Piscataqua-Salmon Falls Hydrologic Unit Code 8 (HUC8) sub-basins (USGS 2021).

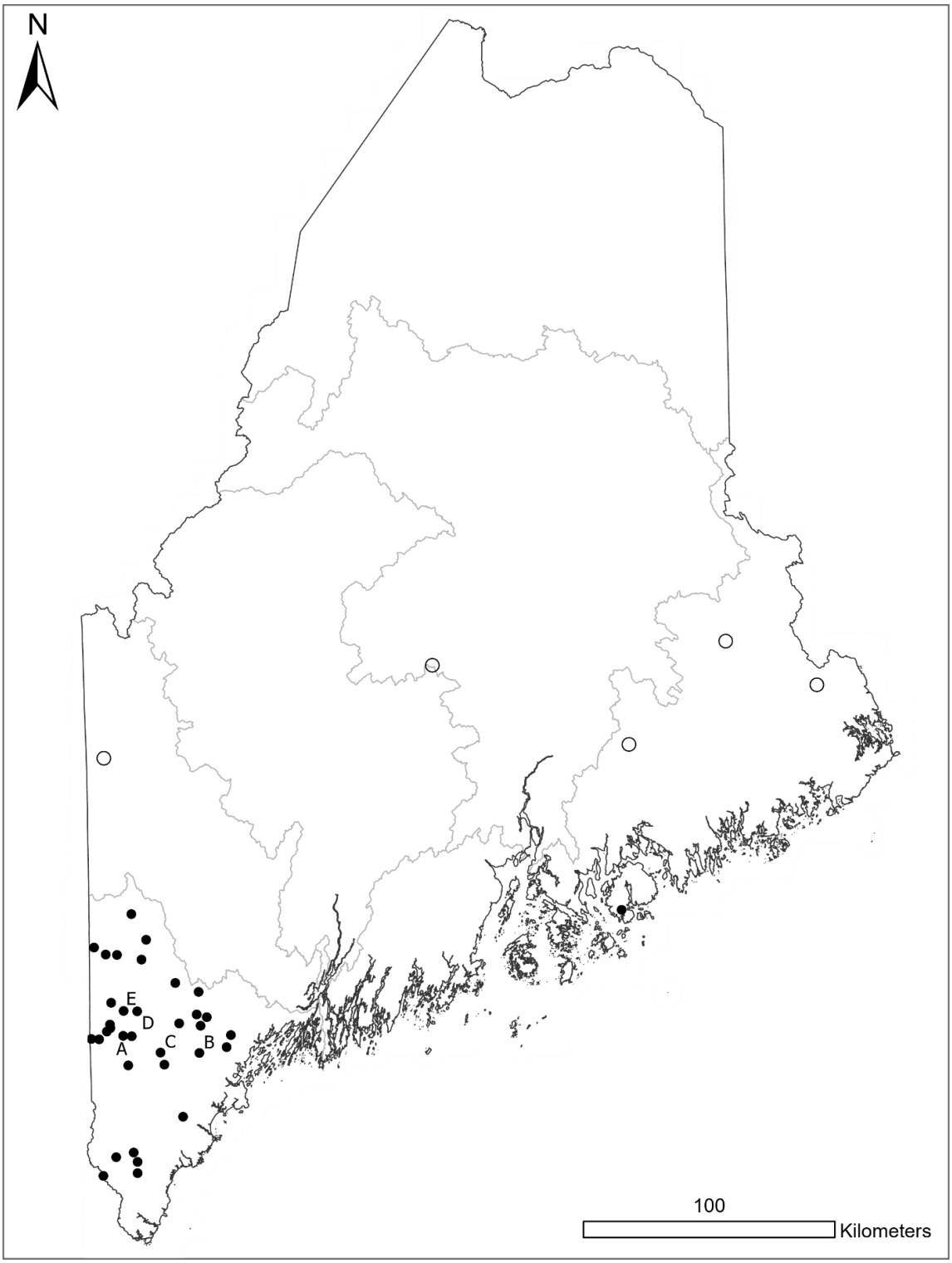


Figure 1.1 Locations of historic bridle shiner records in Maine. Confirmed records are represented by black circles, while likely misidentifications are represented by open circles. River basins (HUC6; USGS 2021) are depicted by light gray lines.

While bridle shiners have not been captured at any of their historically occupied sites in recent years, this may reflect biases in gear, sampled habitat, and survey goals rather than true absences. For example, MDIFW lake and pond surveys between 1950 and 1970 often did not distinguish between minnow species, and sampling equipment used sizes that allowed smaller fish to evade capture (Stone et al. 2001). Therefore, it is possible that bridle shiner populations remain at some or all of these historic sites, but the overall status of bridle shiner populations in Maine is unknown.

We surveyed bridle shiners at 95 locations within 68 waterbodies in southwestern Maine, USA between June 2021 and November 2022. We visited an additional three sites where we were unable to sample with eDNA or a seine net (i.e., could not get permission to sample, $n = 2$; pond dried and could no longer support fish, $n = 1$). Sampling in both years focused on the Saco, Presumpscot, and Piscataqua-Salmon Falls Hydrologic Unit Code 8 (HUC8) sub-basins (USGS 2021).

1.2.2. Site Selection

1.2.2.1. Identification of Historic Sites. We used Google Earth aerial imagery (Google LLC, Mountain View, CA) and in-person site surveys to identify shallow, vegetated habitats within waterbodies where we did not have historic survey coordinates. When river or stream road crossing coordinates were reported in historic records, we used aerial imagery or in-person surveys to search upstream and downstream of the coordinates for vegetated pools. It was not feasible for us to sample larger lakes and ponds in their entirety, so we delineated at least two potential sampling areas (typically the inlet and the outlet) at lakes and ponds. We referred to each historic lake, pond, or stream reach as a "site," and each location where we collected a full sample (2 or 3-L) of water as a "subsite". We chose sampling locations with abundant aquatic

vegetation (when present) that were shallow enough to sample with a 1.2-m high seine net, and split areas into subsites if fish would need to cross a barrier (e.g., expanse of open water or strong current) to move between habitat patches. Each subsite encompassed one patch of contiguous aquatic vegetation. We conducted site visits (n = 47 locations within 32 waterbodies) between June and July of 2021. In 2022, we visited three additional sites with historic bridle shiner records (CRESLK-01/02/03, MARRPD-01/02, SOKOLK-01/02/03; Table 1.1).

Table 1.1 Maine bridle shiner survey sites in 2021-2022 (NAD83 / UTM Zone 19N). We validated 2021 eDNA results with seine net surveys and conducted only eDNA surveys in 2022. We sampled three sites every six weeks between October 2021 and November 2022 (PRESUM-01, OSSIPE, WATBRK). We sampled two sites known to support bridle shiners with three different filter types in 2022 (BARKER, SACONO-03).

Site	Waterbody	Year Sampled	Easting	Northing	Method	Present/ Absent
BARKER	Barker Pond	2021, 2022	359052	4860990	eDNA + seine	**Present**
BEARPD-01	Bear Pond	2021	362597	4890820	eDNA + seine	Absent
BEARPD-02	Bear Pond	2021	363317	4889350	eDNA + seine	Absent
BOOMBR	Unnamed brook	2021	377144	4819870	eDNA + seine	Absent
BROBRK	Brown's Brook	2021	554524	4967970	eDNA + seine	Absent
BURNPD-01	Burnt Meadow Pond	2021	348493	4865680	eDNA + seine	Absent
BURNPD-02	Burnt Meadow Pond	2021	348628	4865060	eDNA + seine	Absent
CROOKR	Crooked River	2021	373952	4873110	eDNA + seine	Absent
GWORKB	Great Works River/ Bauneg Beg Pond	2021	359014	4801930	eDNA + seine	Absent
GWORKN	Great Works River	2021	357524	4805680	eDNA + seine	Absent
GWORKS	Great Works River	2021	358935	4797490	eDNA + seine	**Present**
HIGHLK-02	Highland Lake	2021	358466	4884480	eDNA + seine	Absent
HIGHLK-03	Highland Lake	2021	362817	4879480	eDNA + seine	Absent
HIGHLK-04	Highland Lake	2021	359879	4884760	eDNA + seine	**Present**
JORDAN	Jordan River	2021	382444	4860640	eDNA + seine	Absent
JOSIES	Josie's Brook	2021	369686	4840580	eDNA + seine	Absent
KIMBAL	Kimball Brook	2021	341689	4887230	eDNA + seine	**Present**
LITTLP-01	Little Pond	2021	350779	4884460	eDNA	Unknown*

Table 1.1 Continued.

LITTLP-02	Little Pond	2021	350903	4884350	eDNA	Unknown*
LITTLR	Little River	2021	350489	4803790	eDNA + seine	Absent
MARBRK	Marshall Brook	2021	551554	4902250	eDNA + seine	Absent
OCSACO	Old Course Saco River	2021	346256	4884460	eDNA + seine	Absent
OSSIPE	Ossipee River	2021-2022	353209	4852140	eDNA + seine	**Present**
OSSIPM	Ossipee River	2021	343737	4850670	eDNA + seine	**Present**
PISCAT	Piscataqua River	2021	394376	4847570	eDNA + seine	Absent
PRESUM-01	Presumpscot River	2021-2022	383533	4845090	eDNA + seine	**Present**
PRESUM-02	Presumpscot River	2021	383583	4845200	eDNA + seine	**Present**
PROCPD-01	Proctor Pond	2021	356431	4900500	eDNA + seine	Absent
PROCPD-02	Proctor Pond	2021	356592	4900260	eDNA + seine	Absent
SACONO-01	Saco River	2021	353424	4862040	eDNA + seine	Absent
SACONO-02	Saco River	2021	353642	4862300	eDNA + seine	Absent
SACONO-03	Saco River	2021, 2022	353523	4862370	eDNA + seine	**Present**
SACOSO	Saco River	2021	356715	4851970	eDNA + seine	Absent
SEBAGO-01	Sebago Lake/ Songo River	2021	373567	4863550	eDNA + seine	**Present**
SEBAGO-03	Sebago Lake	2021	370327	4864980	eDNA + seine	Absent
SEBAGO-04	Sebago Lake	2021	375361	4848550	eDNA + seine	Absent
SEBAGO-06	Sebago Lake	2021	381712	4861730	eDNA + seine	Absent
SFALLS	Salmon Falls River	2021	345364	4796430	eDNA + seine	**Present**
SPECPD	Spectacle Ponds	2021	346921	4853590	eDNA + seine	Absent
STANPD-01	Stanley Pond	2021	348533	4854720	eDNA + seine	Absent
STANPD-02	Stanley Pond	2021	348118	4854850	eDNA + seine	Absent
STANPD-03	Stanley Pond	2021	347764	4855830	Seine	Absent†
TRAFPD-01	Trafton Pond	2021	348270	4856280	eDNA + seine	Absent
TRAFPD-02	Trafton Pond	2021	347797	4856810	Seine	Absent†
WATBRK	Unnamed brook	2021-2022	368055	4845460	eDNA + seine	Absent
ANDROS	Androscoggin River	2022	405085	4877020	eDNA	Absent
BRADPD-01	Bradley Pond	2022	351426	4899890	eDNA	Absent
BRADPD-02	Bradley Pond	2022	351071	4899370	eDNA	Absent
BUCKBR	Buck Meadow Brook	2022	350874	4869460	eDNA	**Present**
BUFFBR	Buff Brook	2022	356590	4828550	eDNA	Absent
CANCO	Unnamed pond	.	396733	4837540	None	Absent‡

Table 1.1 Continued.

CARBRK	Carsley Brook	2022	368265	4882340	eDNA	Absent
CHANBR	Chandler Brook	2022	401963	4862410	eDNA	Absent
COLCPD-01	Colcord Pond	2022	342804	4855520	eDNA	**Present**
COLCPD-02	Colcord Pond	2022	342639	4857680	eDNA	Absent
CRESLK-01	Crescent Lake/ Tenny River	2022	382630	4867340	eDNA	**Present**
CRESLK-02	Crescent Lake	2022	383820	4869380	eDNA	**Present**
CRESLK-03	Crescent Lake	2022	382678	4871880	eDNA	Absent
CROOKN	Crooked River	2022	357338	4900570	eDNA	Absent
CROOKS	Crooked River	2022	374475	4870820	eDNA	Absent
DINGLY	Dingley Brook	2022	378789	4863140	eDNA	Absent
DUCKIN	Duck Pond Brook	2022	358274	4884980	eDNA	Absent
DUCKNO	Duck Pond Brook	2022	357422	4889870	eDNA	Absent
EDDYBR	Eddy Brook	2022	393679	4868000	eDNA	Absent
GRTBRK	Great Brook	2022	346438	4846970	eDNA	Absent
HALEY	Unnamed brook	2022	353782	4844520	eDNA	Unknown*
HEATH-01	The Heath	2022	382068	4875170	eDNA	Absent
HEATH-02	The Heath	2022	381920	4874680	eDNA	Absent
INGALS-01	Ingalls Pond	2022	355979	4858140	eDNA	Absent
INGALS-02	Ingalls Pond	2022	356014	4857860	eDNA	Absent
MARRPD-01	Marr Pond	2022	476034	4999620	eDNA	Absent
MARRPD-02	Marr Pond	2022	476743	4999440	eDNA	Absent
MEADBR	Meadow Brook	2022	413718	4869180	eDNA	Absent
MERRIL	Merrill Brook	2022	408744	4855850	eDNA	Absent
MOSQPD	Mosquito Pond	2022	356574	4906770	eDNA	Absent
MUDNO	Mud Pond	2022	358926	4865650	eDNA	**Present**
MUDSO	Mud Pond	2022	348484	4830530	eDNA	Absent
OSSIPR	Ossipee River	2022	353408	4852070	eDNA	Absent
OTTER-01	Otter Ponds (Snake Pond)	2022	378927	4846550	eDNA	Absent
OTTER-02	Otter Ponds (Half Moon Pond)	2022	378321	4846660	eDNA	Absent
PANTHR-01	Panther Pond/ Tenny River	2022	382062	4866260	eDNA	**Present**
PANTHR-02	Panther Pond	2022	381929	4863540	eDNA	Absent
PANTHR-03	Panther Pond	2022	383454	4864880	eDNA	Absent
PISCDN	Piscataqua River	2022	395387	4845110	eDNA	Absent

Table 1.1 Continued.

PISCUP	Piscataqua River	2022	394834	4850480	eDNA	Absent
PRESBG	Presumpscot River	2022	383487	4846940	eDNA	Absent
RACHEL	Unnamed brook	2022	396681	4827270	eDNA	Absent
RANGE-01	Middle Range Pond	2022	389383	4877000	eDNA	Absent
RANGE-02	Middle Range Pond	2022	388929	4874370	eDNA	Absent
REDBRK	Red Brook	2022	391701	4831260	eDNA	Absent
RIDGEB	Unnamed brook	2022	363812	4826610	eDNA	Absent
ROYALR	Royal River	2022	398071	4874310	eDNA	Absent
SHEPR	Shepard's River	2022	345344	4866430	eDNA	Absent
SOKOLK-01	Sokokis Lake	2022	356133	4839980	eDNA	Absent
SOKOLK-02	Sokokis Lake	2022	354643	4841400	eDNA	Absent
SOKOLK-03	Sokokis Lake	2022	354682	4841370	eDNA	Absent
SOPER	Soper Mill Brook	2022	402191	4875470	eDNA	Absent
SYMMES-01	Symmes Pond	2022	348917	4834440	eDNA	Absent
SYMMES-02	Symmes Pond	2022	348864	4834450	eDNA	Absent

*Sample unusable (algae) or lost.
†Received landowner permission to sample post-eDNA surveys. Seine sample only.
‡Pond dry, no water sample taken.

1.2.2.2. eDNA Method Development. We selected three of the historic sites to survey for bridle shiner DNA across all four seasons (map labels A, B, and C, Figure 1.1; OSSIPE, PRESUM-01, and WATBRK; Table 1.1). One site (WATBRK) was a putative null: we did not detect bridle shiners there through either seine net or eDNA survey in 2021, so we did not expect to detect bridle shiner DNA there at any time of year. The second site (OSSIPE) was representative of sites where we detected very few bridle shiners via seine net survey and failed to detect bridle shiner DNA from water samples. We expected eDNA detections to peak during the spawning season at this site. The third site (PRESUM-01) was representative of sites where we successfully detected bridle shiner DNA and captured multiple individuals during a subsequent

seine net survey. We expected to detect bridle shiners with eDNA at this site regardless of season. We collected three replicate, 1-L samples of water from each site approximately every six weeks from 4 October 2021 until 10 November 2022.

In addition to collecting repeated samples at three sites, we selected two sites (map labels D and E, Figure 1.1; BARKER and SACONO-03, Table 1.1) at which to compare three filter pore sizes (2022). We wanted to determine if future studies could use larger filter pore sizes without reducing bridle shiner detection probability as this would save valuable processing time and could potentially allow for filtration in the field. We had captured multiple bridle shiners at these sites in 2021 and expected that these subpopulations would still be present in 2022 and could be reliably detected using eDNA. On 25 September 2022, we collected three groups of five 1-L replicates each at each site: we filtered replicates from Group A using Whatman GF/F 0.7-µm filters (Cytiva, Marlborough, MA), replicates from Group B using Whatman 934-AH™ 1.5-µm filters, and replicates from Group C using Tisch Scientific Grade D 2.7-µm filters (Tisch Scientific, Cleves, OH). We also collected one 1-L field control at each site using the methods described above and filtered the control with a Whatman 0.7-µm filter.

Water collection followed the same strategy as our other 2022 eDNA sampling, except that three 500-mL bottles were held together and submerged as a group to collect each replicate (rather than filling each bottle separately). This ensured that we collected each group of replicates (e.g., A1, B1, and C1) from precisely the same location and at the same time. We used bleached rubber bands to hold the three bottles together and affixed them to a bleached fiberglass pole.

1.2.2.3. Selecting Sites with Unknown Bridle Shiner Presence. In 2022, we endeavored to find additional bridle shiner subpopulations and to sample a variety of habitat types to inform

future species distribution modeling (Chapter 2). While wetlands classified as riverine or lacustrine aquatic bed (Cowardin et al. 1979) by the National Wetlands Inventory (U.S. Fish and Wildlife Service [USFWS] 2022) will likely identify bridle shiner habitat, these classifications are not always available at local scales. To target specific habitats in our bridle shiner surveys, we built a preliminary habitat suitability index model for the state of Maine using ModelBuilder in ArcGIS Pro (version 2.9.1; Esri, Redlands, CA). We included environmental variables such as land cover, catchment position, stream gradient, and wetland type because these variables have been shown to influence bridle shiner occurrence in Connecticut (Jensen and Vokoun 2013; Pregler et al. 2019). Other variables such as substrate type and plant cover are also predictive of bridle shiner habitat but are unavailable in spatial data repositories for water features in Maine. We ranked variable classes from low (1) to high (5) based on habitat data from Jensen and Vokoun (2013) and Pregler et al. (2015, 2019), added the layers together into a final suitability layer, and categorized the raw combined index scores as "least likely," "likely," and "most likely" to support bridle shiner habitat. We ran a binomial generalized linear model (GLM) to determine if suitability category influenced bridle shiner presence. We then calculated the χ^2 analysis of deviance of the model to determine if the habitat suitability covariate significantly improved model fit.

We overlaid a hexagonal grid over the state and used the R package *spsurvey* to select a subset of grid cells to sample (Dumelle et al. 2023, R Core Team 2021). We used a generalized random tessellation stratified (GRTS) survey design to select new water bodies to sample in 2022. The GRTS sampling scheme allowed us to drop grid cells where sites were inaccessible or otherwise unfavorable for sampling while retaining a spatially balanced sampling effort (Brown et al. 2015).

We selected 64 random grid cells within the Saco, Presumpscot, and Lower Androscoggin River watershed units, and chose one waterbody to survey within each cell. We dropped four cells with large lakes that would have required more than three subsites to survey, as our goal was to survey a higher number of waterbodies. We divided the selected waterbodies into three groups based on the preliminary habitat model and planned to survey 20 waterbodies with high predicted habitat suitability, 20 with medium predicted suitability, and 20 with low predicted suitability.

1.2.3. Seine Net Surveys

In 2021, we seined 29 sites that produced a historical (1930s-1940s) or recent (1990s-2010s) record of bridle shiner occurrence. We conducted a power analysis to determine the number of sampling events required to detect a rare fish species (assuming a density < 0.4 fish/sample unit and a detection probability of 50%; Green and Young 1993) and determined that sampling eight locations at each waterbody would provide us with approximately 80% confidence in stating that bridle shiners were not present. Therefore, at each water body, we sampled up to eight locations within the habitat patches where we collected eDNA, as described below.

In Connecticut, bridle shiner young-of-year are large enough to confidently identify beginning in August (Jensen and Vokoun 2013). We began seining in mid-August to avoid misidentifying young bridle shiners and the young-of-year of other minnow species. All fish were handled in accordance with University of Maine Institutional Care and Use Committee (IACUC) protocols (permit number A2021-03-01). Pilot capture efforts at one site showed that young-of-year minnow mortality was high when using a bag seine net, so we only used flat, bagless seines (1.6-mm mesh, 9.1-m length, 1.2-m height) for the remainder of the surveys. We

re-used nets within waterbodies but cleaned and sanitized nets (1% bleach dilution) between unconnected waterbodies to avoid spreading diseases and invasive variable-leaved watermilfoil (*Myriophyllum heterophyllum*).

Because we were primarily interested in bridle shiner presence rather than abundance, we used an adaptive sampling approach (Bonar et al. 1997) and stopped sampling if we captured bridle shiners before the eighth seine sample. We sampled a minimum of three seine samples per site to characterize habitat and fish communities (See Appendix C for community composition methods and results). We fixed one end of the seine in place and dragged the other end to sweep through areas where we either saw minnows or identified patches of submerged or emergent aquatic vegetation. Each full seine drag – covering a surface area of at least 32.8 square meters – was approximately in the shape of a semi-circle and counted as one sample. We then pursed the seine and funneled fish towards the back end of the net, where we collected them and transferred them to an aerated bucket. We counted the number of captured fish in each seine haul and identified them to species, then batch-weighed them by species. We took voucher photos of each species, especially minnow species, using a micro photo tank (8.9 x 3.8 x 3.8 cm). When we caught bridle shiners, we measured up to ten individual fish lengths as well as batch-weighing them. We kept fish contained in buckets until we finished surveying the immediate area so that we did not recount individuals.

1.2.4. Sampling eDNA

In 2021, we began eDNA sampling in early summer to coincide with the start of the bridle shiner spawning period (late May through mid-July in New Hampshire; Harrington 1948). While water levels in May and June were higher due to spring rains, potentially diluting eDNA, the release of fish gametes (and later, the presence of schools of young-of-year fish) should have

increased our probability of capturing bridle shiner DNA during this period (de Souza et al. 2016). In 2022, we ended all eDNA surveys (except for the long-term site surveys) before peak leaf senescence in the fall (mid-October), as humic substances leached from fallen leaves are known to inhibit PCR (Wilson 1997; Eichmiller et al. 2016).

At each sampling location, we collected two 1-L replicates of water (divided between four 250-mL Nalgene bottles) and one 1-L field control following USFWS (2020) protocols. We modified collection protocols slightly in 2022 to include a third 1-L replicate of water at each sampling location to increase the likelihood of capturing bridle shiner DNA in the sample. We also collected each 1-L replicate in two 500-mL Nalgene bottles instead of four 250-mL bottles. We affixed the empty bottles to a 1.2-m PVC or fiberglass sampling pole with rubber bands to collect water. We used separate sampling poles and rubber bands among all replicates and controls at each site and between different sites. We decontaminated sampling poles and other sampling equipment by soaking them in a 10% bleach dilution for ten minutes (Collins et al. 2019; USFWS 2020) and rinsing them with tap water.

We followed several practices to increase our likelihood of detecting bridle shiner DNA and to minimize sample contamination. When sampling streams and rivers, we collected water from areas of low flow such as eddies and backwaters, as DNA in high flow areas may be flushed rapidly downstream (USFWS 2020). As suggested by Goldberg et al. (2018), we collected the sample replicates up to 60-m apart from one another within each subsite to increase the probability of capturing bridle shiner DNA in the sample. Where there was flow, we sampled from downstream to upstream (USFWS 2020; Wood et al. 2021). Some lakes, ponds, and river sites were only accessible by boat or with chest waders: in these instances, we rinsed the outside of a canoe or the waders away from or downstream of the eDNA collection area. When possible,

we collected samples from the shoreline, with only the sampling pole and Nalgene bottle touching the water.

Each water body had a negative field control, which consisted of four 250-mL bottles filled with tap water (USFWS 2020; Wood et al. 2021). Following USFWS (2020) protocols, we opened each field control bottle and exposed the contents to the air for ten seconds. Then, we resealed the bottle and submerged it in the water at the site. We expected that controls would only contain bridle shiner DNA if there was cross-contamination between samples during transport or contamination from the outside of the bottle during filtering.

1.2.4.1. Laboratory Methods. We transported water samples on ice to the University of Maine eDNA CORE lab, where we stored them at 4°C until filtration within approximately 24 hours of collection (Hinlo et al. 2017). We began the summer by using Whatman 7190-004 1.0-µm, 47-mm diameter cellulose nitrate filters (Cytiva, Marlborough, MA) but switched to Whatman GF/F 0.7-µm, 47-mm diameter glass fiber filters on 29 June 2021 (samples BURNPD-01/02, SPECPD, OSSIPE, JORDAN, PRESUM-01/02, SACONO-01/02, BARKER and MARBRK were filtered with cellulose nitrate filters; Table 1.1). Higher temperatures at the sites facilitated the growth of bacteria and algae, which quickly clogged the cellulose nitrate filters. We vacuum-filtered as much of each 1-L replicate as possible through a 0.7-µm glass fiber filter to isolate the target DNA from the sample (Lacoursière-Roussel et al. 2016a; Hinlo et al. 2017; Goldberg et al. 2018; Plough et al. 2018). When necessary, we used up to three filters per replicate and combined them prior to quantitative PCR (qPCR). We stored the filters in a -20°C freezer, and eDNA CORE lab personnel extracted DNA from the filters using a modified DNeasy Blood & Tissue kit and associated Standard Operating Procedure (SOP; Appendix A).

We designed a species-specific bridle shiner qPCR primer-probe set based on the TaqMan MGB-NFQ chemistry, which amplifies a 149bp (base pair) portion of the cytochrome b (*Cytb*) gene (See Appendix B for primer development methods). We performed qPCR on a Bio-Rad CFX96 Real-Time System thermal cycler (Bio-Rad Laboratories, Hercules, CA) using reaction chemistry: 10-µL TaqMan Environmental Master Mix 2.0, 3-µL template DNA, assay concentrations of 1-µM primers, 500nM probe, and nuclease-free water to bring the reaction volume to 20-µL. The thermal protocol for all qPCR reactions was as follows: 95°C for 10 minutes followed by 50 cycles of 15 seconds at 95°C and 15 seconds at 60°C. Possible PCR inhibition was removed from samples using Zymo OneStep PCR Inhibitor Removal Kits (Zymo Research, Irvine, CA, USA) using the manufacturer's protocol.

We ran four technical replicates per eDNA sample, including field negative controls and PCR negative controls. We prepared the latter by using DNA-free water in place of extracted sample template when plating. We also included positive controls in the form of synthetic gene (gBlock) fragments corresponding to the *Cytb* gene region of our eDNA assay. These positive controls included 18 reactions across 13 well plates at concentrations of 10, 50, 250, 1,250, 6,250, and 31,250 copies per µL (Wood et al. 2020) to provide a qPCR standard calibration curve (York 2016).

Bio-Rad CFX Manager software was used to estimate *Cq* values from qPCR fluorescence curves. We considered a *Cq* value below 45 to be a positive bridle shiner DNA detection (Wilcox et al. 2013). We log-transformed the initial gBlock dilution series concentrations and plotted average *Cq* values at each concentration and used the resulting linear regression equation to estimate the initial DNA concentrations for each positive qPCR replicate.

1.2.5. Statistical Analyses

1.2.5.1. Historic Sites: Comparing Seine Net and eDNA Detection Probability. Survey methods such as eDNA inherently have imperfect detection: the collected water sample may not capture target DNA from the environment, and DNA extraction and amplification techniques may not detect small quantities of DNA present in the sample (Dorazio and Erickson 2018; Mize et al. 2019). Occupancy models infer species occupancy while accounting for imperfect detection (MacKenzie et al. 2002; Lahoz-Monfort et al. 2016). We conducted two single-season occupancy models in the R (version 4.2.2; R Core Team 2022) package *unmarked* (version 1.2.5; Fiske and Chandler 2011) to compare seine net and eDNA detection probabilities. For these analyses, we only compared sites which we surveyed via both seine netting and eDNA sample collection (n = 29 sites). Seine net surveys can be divided into two levels of sampling: (1) fish presence within a waterbody and (2) fish detection within a seine net haul. To compare seine and eDNA surveys directly, we aggregated water samples by waterbody so that each waterbody had a maximum of eight seine net hauls and eight 1-L eDNA samples (from up to four subsites). We aggregated the results of the four technical PCR replicates from each eDNA sample so that we were only modeling the water sampling process and not the qPCR replicate detection probability, which we modeled in later analyses (Schmidt et al. 2013). Thus, we only modeled occupancy at the waterbody scale and not the subsite scale.

We performed exploratory occupancy analyses using site area and upstream drainage area as covariates, but the null models were favored in all instances. As we were only comparing the 29 sites sampled in 2021, most models with covariates using this reduced dataset did not converge. This was determined by a warning message from *unmarked* stating that an individual model had not converged, likely because the standard errors of the coefficients were high. We

back-transformed coefficient estimates for each null model to obtain the mean detection probability (p) and occupancy probability (ψ), along with their associated standard errors. Using the mean detection probability and standard error, we calculated the cumulative probability (P^*) of detecting bridle shiner DNA in *N* samples:

$$P^* = 1 - (1 - p)^N$$ (McArdle 1990).

These analyses provided us with the probability of detecting bridle shiners at a site and, conversely, the probability of failing to detect bridle shiners at a site, given that they were present.

1.2.5.2. eDNA Sampling Optimization. Our bridle shiner eDNA surveys inherently included three levels of sampling: (1) fish eDNA presence within a waterbody, (2) DNA presence within a sample, and (3) DNA detection within replicate subsamples. These data can be analyzed using multiscale occupancy models (Nichols et al. 2008; Dorazio and Erickson 2018) to refine future sampling techniques by determining the sample volume and number of replicates needed to achieve a threshold detection value (Dorazio and Erickson 2018).

We fit the three-level occupancy model developed by Nichols et al. (2008) and Mordecai et al. (2011) using WinBUGS (Kéry and Royle 2016) to determine the occupancy, availability, and detection processes underlying our eDNA data (Schmidt et al. 2013; Appendix D). For this analysis, we used the first eDNA survey from each of our 2021 and 2022 sites and did not include any additional surveys from sites visited more than once (e.g., filter pore size or seasonal comparison sites). We conducted these analyses at the subsite (habitat patch) scale rather than at the whole waterbody scale because we could not survey lakes and ponds in their entirety. The habitat patch scale is also more biologically relevant, as patches were separated from one another by barriers to bridle shiner movement.

The multiscale model consists of three coupled Bernoulli trials (Schmidt et al. 2013):

$$z_i \sim Bernoulli(\psi_i),$$

$$a_{ij}|z_i \sim Bernoulli(z_i\theta_{ij}), \text{ and}$$

$$y_{ijk}|a_{ij} \sim Bernoulli(a_{ij}p_{ijk})$$

The data (y_{ijk}) are the binary indicators of detection and non-detection of bridle shiner DNA at site *i* (*i* = 1,…,93), in water sample *j* (*j* = 1,…,5), and in qPCR technical replicate *k* (*k* = 1,…,4; Schmidt et al. 2013). Bridle shiners could be either present ($z_i = 1$) or absent ($z_i = 0$) from a subsite, and bridle shiner DNA could be either present ($a_{ij} = 1$) or absent ($a_{ij} = 0$) from a water sample taken at subsite *i*. Given that bridle shiners were present at subsite *i* and bridle shiner DNA was captured in water sample *j*, p_{ijk} is the probability of detecting the DNA in qPCR replicate *k*. We used vague uniform priors for all model parameters as outlined by Kéry and Royle (2016):

$$\psi \sim \text{dunif}(0,1),$$

$$\theta_j \sim \text{dunif}(0,1), \text{ and}$$

$$p_k \sim \text{dunif}(0,1)$$

We set the initial value of ψ at 0.5 (Kéry and Royle 2016). We derived the total number of occupied sites ($\sum z_i$), the total number of samples with presence ($\sum a_{ij}$), the mean detection probability ($\hat{p}$), and the mean availability probability ($\hat{\theta}$). We ran a total of 25,000 model iterations and discarded the first 2,000 as burn-in. Finally, we used the mean detection probability ($\hat{p}$) to calculate the cumulative probability (P^*) of detecting bridle shiner DNA after *k* qPCR replicates:

$P^* = 1 - (1 - \hat{p})^k$ (McArdle 1990; Schmidt et al. 2013).

We repeated this process with the mean availability probability ($\hat{\theta}$) to determine the cumulative probability (θ^*) of detecting bridle shiner DNA in *j* water samples:

$$\theta^* = 1 - (1 - \hat{\theta})^j$$ (McArdle 1990; Schmidt et al. 2013).

1.2.5.3. eDNA Method Development

1.2.5.3.1. Filter Pore Size. We conditioned a second hierarchical occupancy model using only sites with known presence (*n* = 17 sites with a total of *n* = 38 site visits). We determined that the cumulative probability of detecting bridle shiner DNA in two or three water samples was high enough to examine differences in filter pore size detection and seasonal patterns of bridle shiner eDNA availability using linear regressions rather than occupancy models. We e ran a series of binomial GLMs using DNA detection (1) and non-detection (0) as the response variable and site, pore size, sample replicate, and qPCR replicate as covariates. We performed model selection using the *glmulti* package (version 1.0.8; Calcagno 2020), which evaluates the Akaike Information Criterion (AIC; Akaike 1974) corrected for small sample sizes (AICc) for each linear combination of model covariates. We then performed a χ^2 analysis of deviance on the top model to determine the significance of the retained covariates to bridle shiner presence.

1.2.5.3.2. Seasonality of eDNA Sampling. To examine seasonal variability in bridle shiner DNA detection, we ran two sets of binomial GLMs: one set with all eDNA sites and one set including only the three year-round sites OSSIPE, PRESUM-01, and WATBRK. We included month, year, season ("fall," "spring," "summer," "winter"), standardized day of year (1 January = 1, 31 December = 365), day of year squared, and period ("active" or "winter") as covariates and tested all linear and quadratic combinations using *glmulti*. We considered the beginning of the spawning period (late May) through peak leaf drop (mid-October) to be the "active" period,

and the remainder of the year to be the "winter" period. We then calculated a χ^2 analysis of deviance of the two models (all sites and long-term sites only) with the lowest AICc.

1.3. Results

1.3.1. eDNA Assay Performance

The qPCR primer-probe set successfully amplified synthetic and *in vitro* bridle shiner DNA targets. The assay did not amplify any off-target species' DNA (Appendix B). Serial dilutions of gBlock fragments confirmed that our qPCR assay detected eDNA at the lowest test concentration of 10 copies per reaction (average *Cq* = 36.2; Figure 1.2). The PCR efficiency was high (108.5%). We did not encounter any false positives when testing the field and laboratory controls.

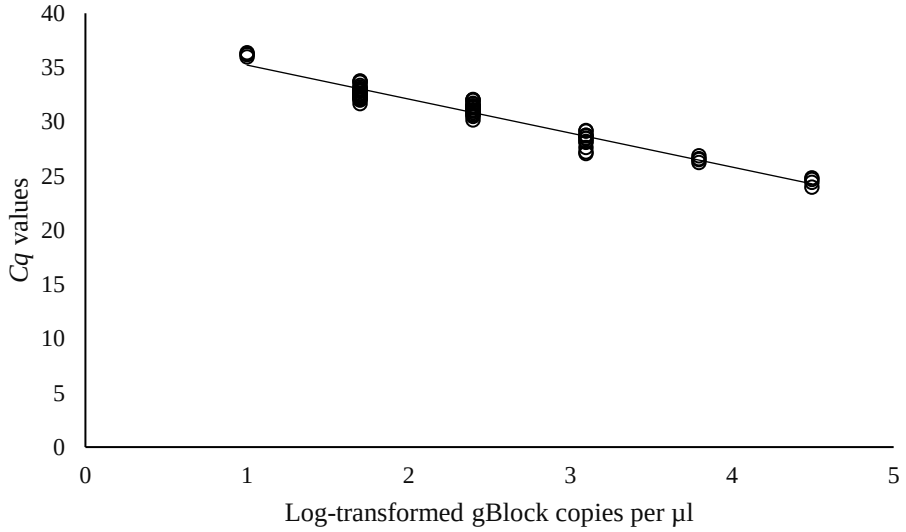


Figure 1.2 Log-transformed gBlock dilution series and corresponding *Cq* values. The x-intercept is estimated to be 38.4 log-copies and the slope of the line is -3.13. $R^2 = 0.95$.

1.3.2. Historic Sites: Comparing Seine Net and eDNA Detection Probability

We sampled water at 50 locations within 33 sites with historic bridle shiner presence between 1 June and 16 July 2021 (*n* = 30 waterbodies) and between 9 July and 11 October 2022

(*n* = 3 waterbodies; Table 1.1). We conducted seine net surveys at 29 of the 30 sites sampled in 2021 but were unable to seine in 2022. We detected bridle shiners or their DNA at 11 historically occupied sites and did not detect them with either method at 21 sites. We were unable to process water samples from one pond (LITTLP-01/02; Table 1.1).

All but one *Cq* value fell below the 45-cycle threshold: one sample from KIMBAL (water sample replicate 2, qPCR technical replicate 4) had a *Cq* of 48.34, so we ran the sample a second time and observed a *Cq* value of 45.65 (Table 1.2). We considered this to be a positive detection because of the more definite amplification and because a qPCR replicate from the first water sample replicate also amplified (*Cq* = 39.43). The historic site detection with the lowest *Cq* value (*Cq* = 37.06) came from HIGHLK-04 (Table 1.2).

Table 1.2 *Cq* values of all positive detections of bridle shiner eDNA in 2021 and 2022 by water sample replicate and qPCR technical replicate. An A, B, or C preceding the replicate number denotes samples taken for the filter pore size comparison in one of three sizes. All other replicates were filtered with a 0.7µm pore size.

			Cq value			
Collection Date	Site	Replicate	pcr1	pcr2	pcr3	pcr4
6/14/2021	PRESUM-01*	R1 (1st run)	.	.	38.91	41.09
6/14/2021	PRESUM-01*	R1 (2nd run)	39.31	.	40.00	.
6/14/2021	PRESUM-02*	R1 (1st run)	.	41.37	.	.
6/14/2021	PRESUM-02*	R1 (2nd run)	39.00	38.77	.	38.65
6/16/2021	BARKER	R1	39.30	39.22	40.04	40.28
6/16/2021	BARKER	R2	39.15	38.36	40.07	38.46
6/30/2021	GWORKS	R2	40.36	.	.	.
7/6/2021	KIMBAL	R1	.	.	39.43	.
7/6/2021	KIMBAL†	R2 (1st run)	.	.	.	48.34
7/6/2021	KIMBAL†	R2 (2nd run)	.	45.65	.	.
7/8/2021	HIGHLK-04	R1	37.06	.	37.09	.
7/14/2021	SEBAGO-01	R2	.	.	.	37.91
7/16/2021	OSSIPM	R1	40.73	.	.	41.07
6/16/2022	OSSIPE	R2	.	.	.	38.18
6/16/2022	OSSIPE	R3	.	38.63	41.86	.
6/16/2022	PRESUM-01	R1	.	38.04	38.62	.

Table 1.2 Continued.

6/16/2022	PRESUM-01	R2	.	37.92	35.82	.
7/15/2022	COLCPD-01	R1	.	38.46	.	.
8/3/2022	PANTHR-01	R1	36.21	.	38.09	.
8/3/2022	PANTHR-01	R2	.	.	.	41.07
8/3/2022	PANTHR-01	R3	.	36.20	.	40.24
8/15/2022	BUCKBR	R1	32.50	31.76	29.68	31.75
8/15/2022	BUCKBR	R2	38.20	36.13	35.06	42.61
8/15/2022	BUCKBR	R3	37.77	36.39	34.07	40.13
8/15/2022	MUDNO	R2	38.23	.	38.36	.
9/25/2022	BARKER	A-R1	.	37.54	.	.
9/25/2022	BARKER	A-R2	.	37.22	38.62	.
9/25/2022	BARKER	A-R3	.	36.28	.	35.49
9/25/2022	BARKER	A-R4	.	38.29	.	36.87
9/25/2022	BARKER	A-R5	.	.	.	38.32
9/25/2022	BARKER	B-R1	.	36.50	40.71	.
9/25/2022	BARKER	B-R2	35.08	39.56	.	.
9/25/2022	BARKER	B-R4	.	39.50	.	40.39
9/25/2022	BARKER	B-R5	.	37.93	39.46	.
9/25/2022	BARKER	C-R2	35.75	.	37.44	40.17
9/25/2022	BARKER	C-R3	.	.	36.44	.
9/25/2022	BARKER	C-R4	.	38.37	37.20	40.22
9/25/2022	SACONO-03	A-R4	.	38.86	.	.
9/25/2022	SACONO-03	A-R5	.	37.77	.	.
9/25/2022	SACONO-03	B-R1	35.60	.	.	39.00
9/25/2022	SACONO-03	B-R5	36.87	.	38.66	.
9/25/2022	SACONO-03	C-R3	.	.	38.70	.
9/25/2022	SACONO-03	C-R5	.	.	.	41.50
10/5/2022	CRESLK-01	R1	34.82	38.65	34.04	37.03
10/5/2022	CRESLK-01	R2	.	37.94	37.63	39.35
10/5/2022	CRESLK-02	R2	38.22	38.85	37.90	37.39
10/5/2022	CRESLK-02	R3	39.48	42.84	.	39.50

Note: *Used these samples from known positive sites to evaluate the bridle shiner assay. Following late-cycle amplifications, we used a higher starting template amount to rerun these samples and all subsequent samples.
† Originally run and pcr4 had a late amplification (48.34). Reran and pcr2 had a more definite amplification (45.65), considered positive.

We identified bridle shiner DNA in seven water bodies in 2021 and captured the fish at six of these (BARKER, GWORKS, HIGHLK-04, KIMBAL, OSSIPM, PRESUM-01/02, and SEBAGO-01; Figure 1.3). We also captured bridle shiners at two sites where we had not

detected them with eDNA (OSSIPE and SFALLS) and in one location where we did not collect eDNA samples (SACONO-03). There was one site (GWORKS; Figure 1.3) where we detected DNA but did not capture any bridle shiners with the seine. Overall, we captured bridle shiners at nine sites using seine nets, and at six sites using both methods (Table 1.1). In 2022, we detected bridle shiner DNA at two locations within one historic site (CRESLK-01 and CRESLK-02; Figure 1.3) and at two sites where we had not detected (OSSIPE) or had not collected (SACONO-03) DNA in 2021. As these eDNA surveys were not accompanied by seine net surveys, we excluded them from the single-season occupancy models.

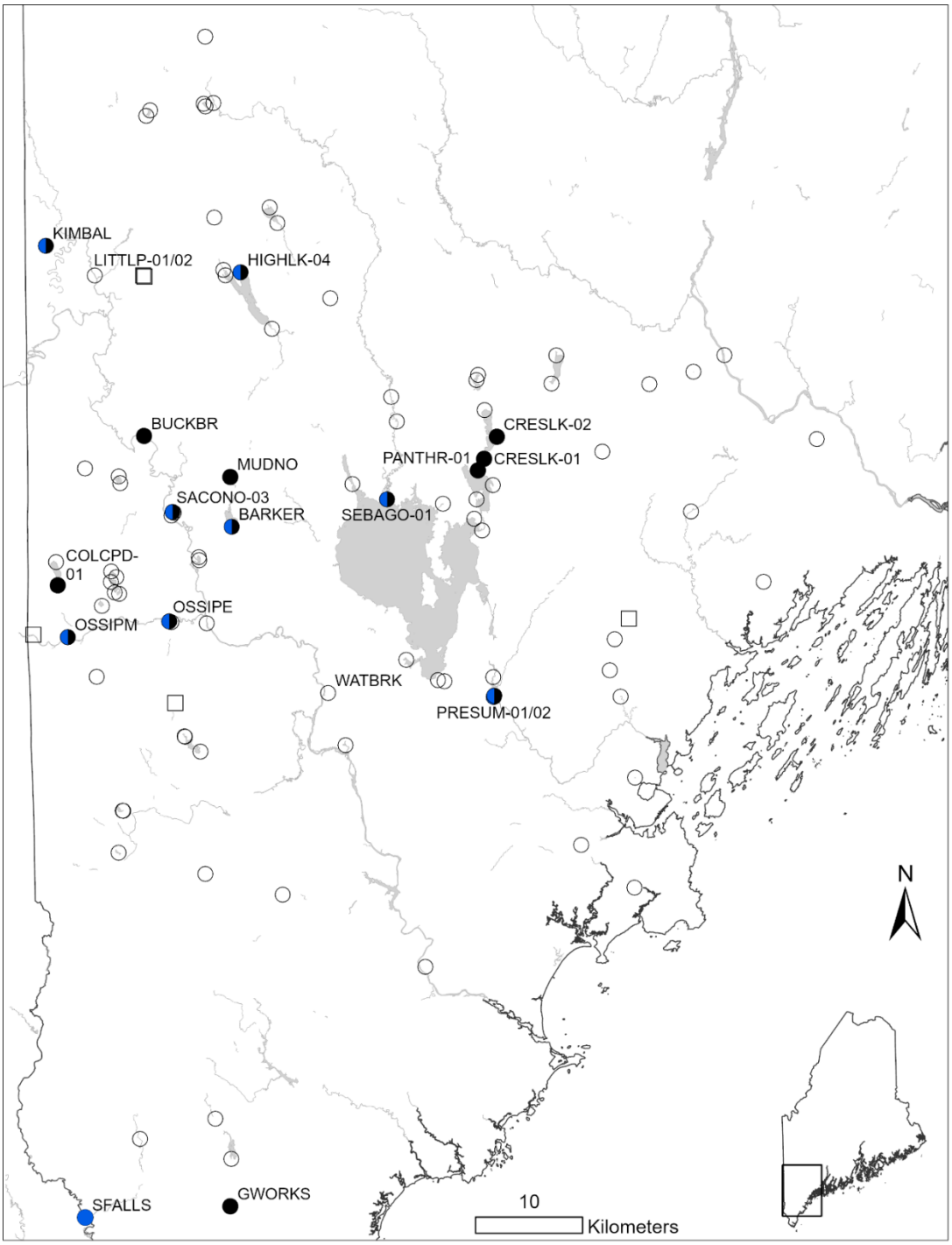


Figure 1.3 We conducted bridle shiner surveys of historically occupied and new sites in southwestern Maine using seine netting (2021) and eDNA (2021-2022). We detected bridle shiner subpopulations using only eDNA (black circles, *n* = 7), only seine netting (blue circle, *n* = 1), or with both methods (blue and black circles, *n* = 9). We were unable to conduct surveys at two historic sites and were unable to process samples from four additional locations (open squares, *n* = 6). We failed to detect bridle shiners with one or both survey methods at 78 sites (open circles).

We estimated the detection probability of a single seine net haul to be $p = 0.20 \pm 0.08$. Our cumulative detection probability over eight hauls was $P^* = 0.83 \pm 0.14$. We collected between two and eight 1-L eDNA replicates per site (mean = 2.7), depending on the size of the site and the number of habitat patches. There were seven sites with at least one eDNA detection. This translated to a detection probability of $p = 0.25 \pm 0.12$ per 1-L sample replicate. The cumulative detection probability over eight sample replicates was $P^* = 0.90 \pm 0.09$.

The seine net occupancy model estimated that 11 (38.0 ± 12.3%) of the 29 seined sites were occupied, and the eDNA occupancy model estimated that 14 (48.8 ± 23.7%) of these sites were occupied. The standard error of these estimates was high because the number of seine samples per site varied (detection error of 51.3% with three seine hauls) and because we collected a minimum of two eDNA replicates per site (detection error of 56.9% with two eDNA replicates). Our observed occupancy was 11 (37.9%) out of 29 modeled sites, which matched the seine net model estimate.

We determined that we would need eight seine net hauls (Figure 1.4a) or six 1-L sample replicates (Figure 1.4b) per waterbody to obtain cumulative detection probabilities greater than $P^* = 0.80$. We would need 11 1-L eDNA samples and 14 seine net hauls per waterbody to obtain cumulative detection probabilities greater than $P^* = 0.95$.

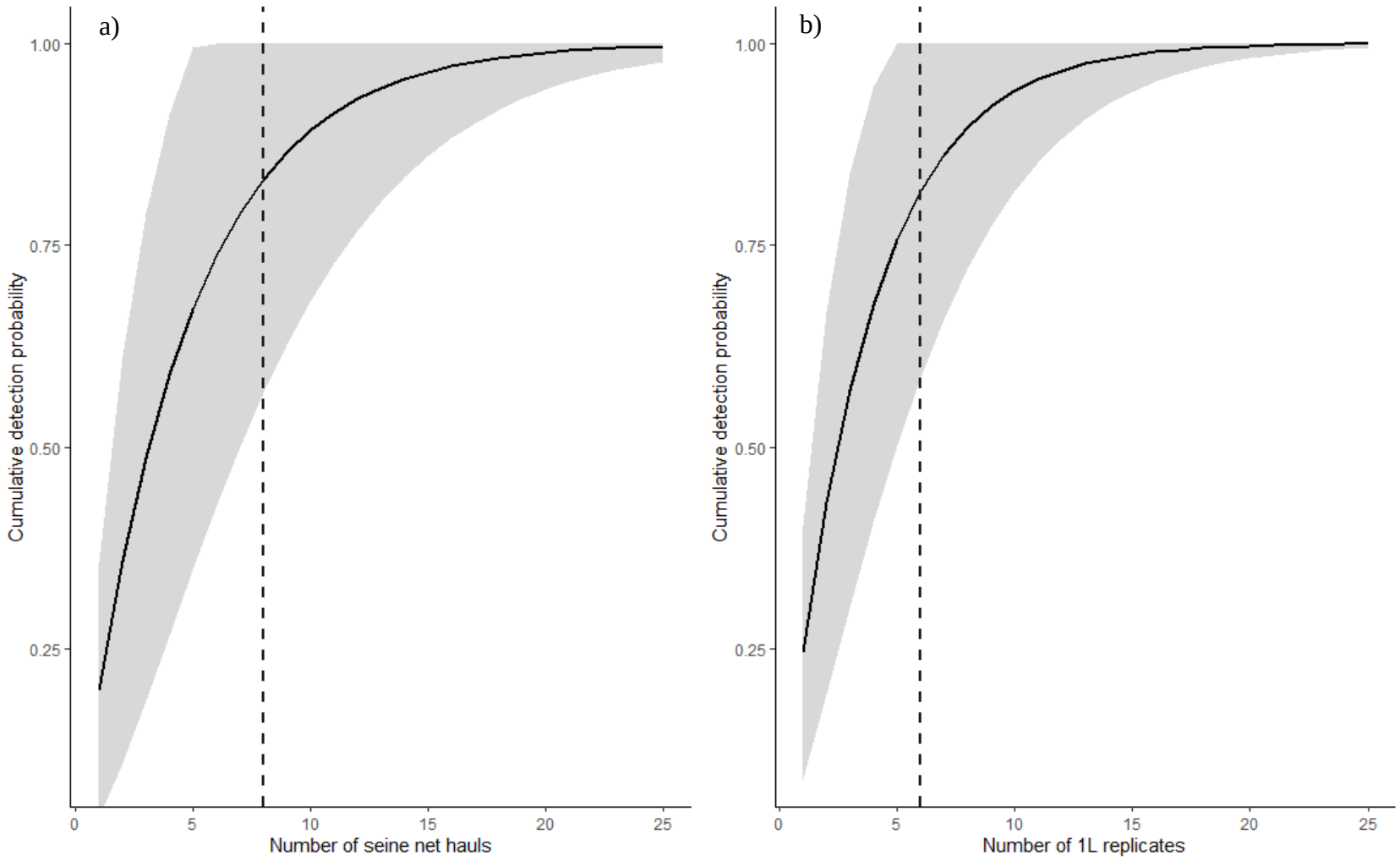


Figure 1.4 Cumulative probability of bridle shiner detection (P^*) over 25 sample replicates for a) seine netting and b) eDNA at the waterbody scale.

While we spread out our seine net hauls across an entire waterbody (e.g., eight seine samples per lake), we took one eDNA sample from each subsite or habitat patch within a waterbody. If we had taken an equivalent number of seine net and eDNA samples at each waterbody, the eDNA detection probability would have been higher than that of seine netting ($P^* = 0.90$ for eight eDNA samples or 8-L of water), but we only collected one sample (2-L total) per habitat patch within a site. Therefore, at sites with three subsites, the probability of detection with eDNA was 81.6%, but at sites with two subsites the probability of detection was only 67.7%.

1.3.3. eDNA Sampling Optimization

We detected bridle shiners at 17 total locations across both years of the study and did not detect them at 76 locations (Figure 1.3). *Cq* values across all samples ranged from 29.68 (BUCKBR) to 48.34 (KIMBAL), with only the sample from KIMBAL passing the 45-cycle threshold (Table 1.2). The preliminary habitat model correctly assigned high habitat suitability index values to several waterbodies currently occupied by bridle shiners and to areas where habitat was abundant, but no bridle shiners were detected. The model also correctly assigned lower values to sites where bridle shiners were once reported, but which currently do not support bridle shiners or their habitat. It is possible that the habitat in these areas has changed since the original bridle shiner sightings in the 1930s, or that bridle shiners never occupied those sites to begin with. In 2022, we surveyed 19 locations with high habitat suitability index values, 26 locations with medium index values, and 16 locations with low index values (57 total locations across 39 waterbodies). We detected bridle shiners with eDNA at 10 of these locations, including in four waterbodies where they were previously undocumented (BUCKBR, COLCPD-01, MUDNO, and PANTHR-01; Table 1.1, Figure 1.2). While the suitability indices appeared to

reliably distinguish between areas with suitable and unsuitable habitat, differences between the three model categories (high, medium, and low predicted suitability) were not statistically significant ($p > 0.05$). The χ^2 analysis of deviance of the model showed that habitat suitability did not significantly improve model fit ($p = 0.36$).

We found that the mean availability probability at the water sample level was $\hat{\theta} = 0.58$ (95% CRI: 0.39, 0.74), and the mean detection probability at the qPCR replicate level was $\hat{p} = 0.57$ (95% CRI: 0.46, 0.67). The cumulative detection probability (P^*) of all four qPCR replicates was $P^* = 0.97 \pm 0.02$. The cumulative availability probability (θ^*) with two water sample replicates (2021 samples) was $\theta^* = 0.82 \pm 0.08$. With three water sample replicates (summer 2022 samples), the cumulative availability probability increased to $\theta^* = 0.92 \pm 0.05$. We plotted the cumulative detection (P^*) and availability (θ^*) probabilities and determined that we would need two qPCR technical replicates and two 1-L water sample replicates to obtain cumulative probabilities greater than 0.80, and four qPCR technical replicates (Figure 1.5a) and four 1-L water sample replicates (Figure 1.5b) to obtain cumulative probabilities greater than 0.95.

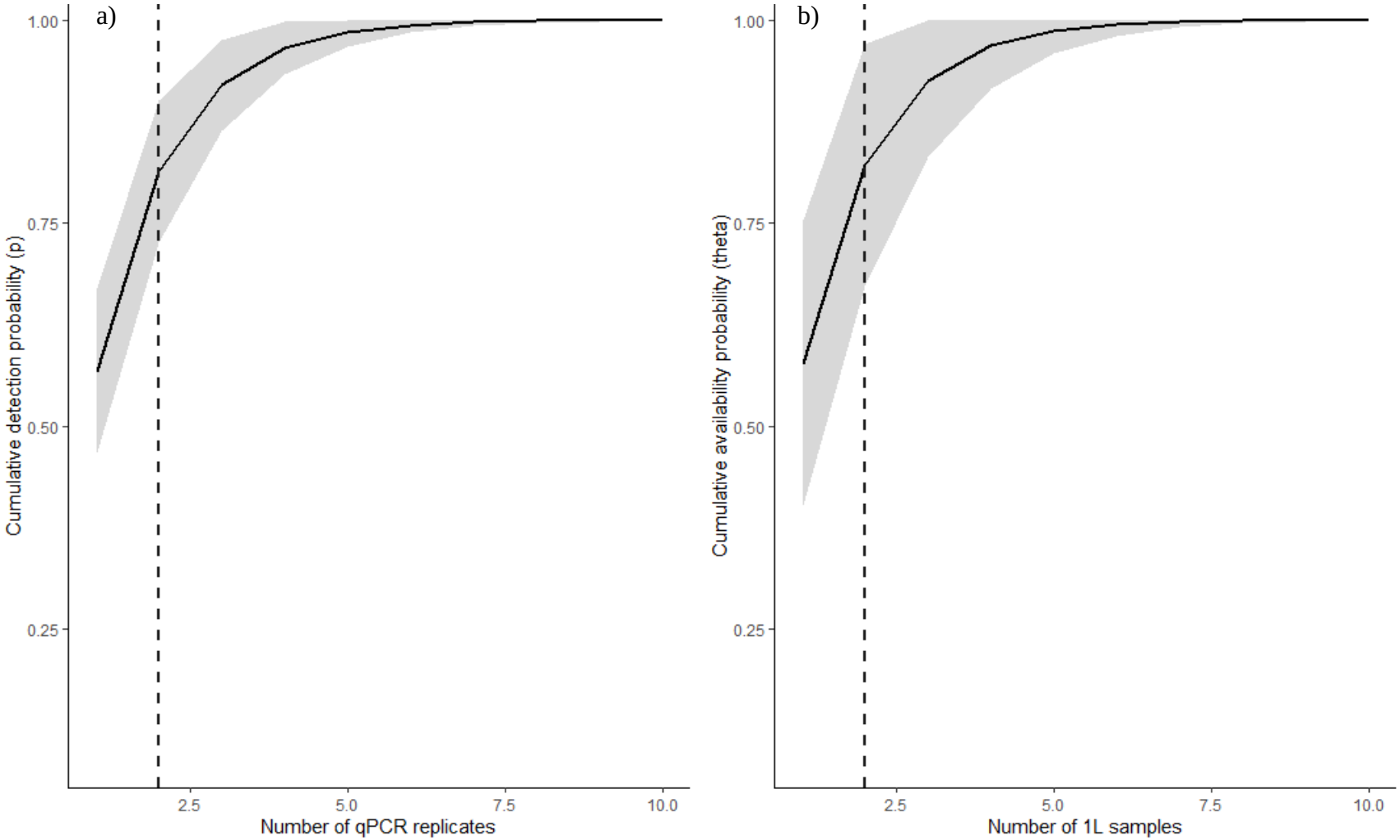


Figure 1.5 Cumulative probability of a) bridle shiner DNA detection over 10 qPCR replicates (P^*) and b) cumulative probability (θ^*) of bridle shiner DNA availability over 10 (1-L) water sample replicates.

Overall, these results suggest that our eDNA survey effort gave us sufficient power to detect bridle shiners at the subsite (or habitat patch) scale. We detected bridle shiners at 17 out of 93 subsites (18.3%), and the model predicted that 19 of the subsites were occupied ($\psi = 0.20$, 95% CRI: 0.12, 0.32). The majority of our eDNA non-detections are therefore likely to reflect true absences. We were unable to detect bridle shiners at 21 of 32 (65.6%) historic sites, suggesting that bridle shiners may have become extirpated from a sizable portion of their historic range in Maine.

1.3.4. eDNA Method Development

1.3.4.1. Filter Pore Size. We did not detect a significant difference in bridle shiner DNA presence between filter pore sizes. This is likely a reflection of sample size, bridle shiner abundance, and eDNA distribution at the sites. In 2021, we captured 11 bridle shiners at BARKER and 29 at SACONO-03. eDNA detection was high at BARKER in 2021, with all four qPCR replicates in both water samples containing bridle shiner DNA.

We found only minor differences in eDNA detection between filter pore sizes. In 2022, we detected bridle shiner DNA in 70% of water samples filtered with 0.7-µm filters, in 60% of water samples filtered with 1.5-µm filters, and in 50% of water samples filtered with 2.7-µm filters. We detected eDNA more reliably at BARKER than at SACONO-03, with 12 positive water samples at BARKER and six positives at SACONO-03. Fish may have been more concentrated in the middle and on the eastern end of the SACONO-03 site, as water samples R3 and R5 (eastern end of pond) were consistently positive while samples R1 and R2 (western end of pond) only produced one positive detection between the three filter sizes.

We fit a binomial GLM and found that only site significantly affected the presence of DNA in the sample ($p = 0.003$). Across filter sizes and sites, qPCR replicate 2 was significantly

more likely to contain bridle shiner DNA ($p = 0.04$) than other qPCR replicates. However, when we compared covariate combinations in *glmulti*, the most supported model did not include qPCR replicate as a covariate. The χ^2 analysis of deviance of the full model revealed that including site in the model significantly improved model fit ($p = 0.001$), but that filter pore size did not ($p = 0.74$).

1.3.4.2. Seasonality of eDNA Sampling. We surveyed three sites (OSSIPE, PRESUM-01, and WATBRK) eleven times between 10 June 2021 and 10 November 2022. We detected bridle shiner DNA on only two of the eleven sampling occasions: twice at PRESUM-01 (14 June 2021 and 16 June 2022), where bridle shiners were abundant in 2021, and once at OSSIPE, where we only caught one bridle shiner in 2021. This positive detection on 17 June 2022 coincided with the presence of many young-of-year fish of at least two species: a cyprinid species (likely bridle shiner) and a sucker species. We did not detect bridle shiner DNA at any time of year at WATBRK, corroborating the results of our seine net surveys and suggesting that bridle shiners were truly absent from the site.

Among all sites, our earliest detection of bridle shiner DNA was 14 June 2021, and our latest detection was 5 October 2022 (Figure 1.6). We found that a quadratic effect on sampling date was best supported ($p = 0.002$), with considerably less support for other seasonal variables. We were more likely to detect bridle shiners when sampling closer to the average sampling day (24 July) rather than earlier or later in the year.

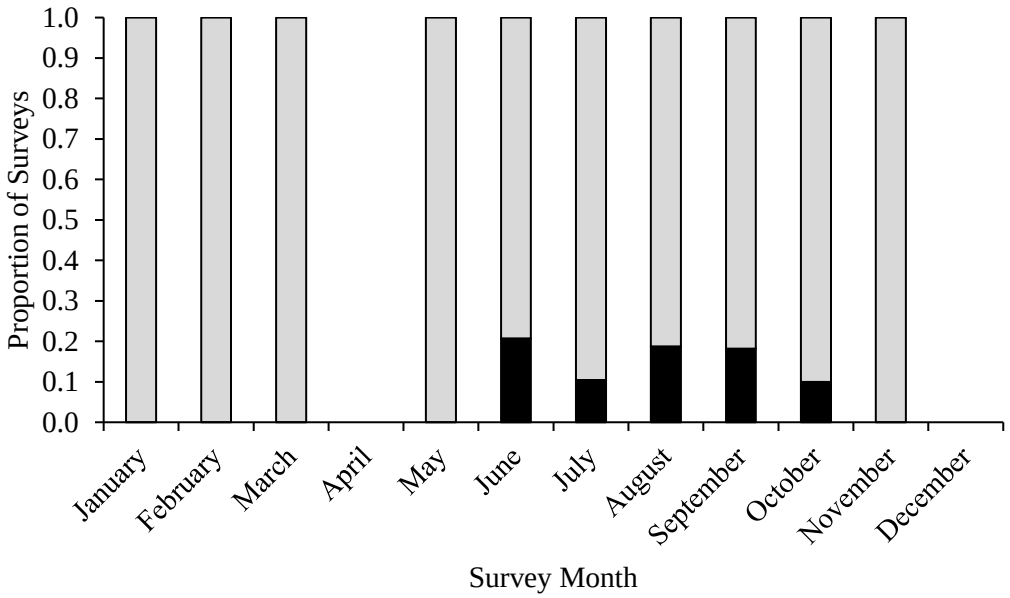


Figure 1.6 Proportion of positive (black) and negative (gray) bridle shiner eDNA surveys by month (2021-2022). Historic site, new/unknown site, long-term, and filter pore size surveys are all included.

We repeated these steps using eDNA survey results from only the three long-term sites. The top model, as determined by *glmulti*, retained sampling day squared but not season. A χ^2 analysis of deviance revealed that the quadratic effect on sampling date significantly improved model fit ($p = 0.001$) over the null model. This model showed the same relationship between bridle shiner presence and sampling day squared as above: we were more likely to detect bridle shiners when sampling closer to the average sampling day (29 June) at long-term sites.

1.4. Discussion

This study demonstrated the viability of both eDNA and seine netting for detecting and monitoring bridle shiners in Maine. We developed and tested a targeted primer-probe assay for bridle shiner and found that it was capable of distinguishing bridle shiner DNA from closely related, sympatric species (Appendix B). We confirmed remnant bridle shiner subpopulations at 11 of 32 historically occupied sites using eDNA and traditional fisheries techniques, then

documented their presence in four additional waterbodies using eDNA (Figure 1.3). We determined that bridle shiners can be detected using eDNA methods between June and October in Maine, but that surveys in early to midsummer have a higher likelihood of detecting this species (Figure 1.6). We also successfully detected bridle shiner DNA with three filter pore sizes, which could reduce sample processing time and cost in future surveys.

Our eDNA and seine net survey efforts gave us sufficient power (>0.80) to detect bridle shiners at both the patch and waterbody scale. Therefore, most of our non-detections likely reflected true absences of bridle shiners at a site. This suggests that bridle shiners have become extirpated from approximately 65% of their historically occupied sites, and points to a significant decline of the species in Maine. These findings are consistent with the declines reported in other regions of the bridle shiner's range (Kilian et al. 2011; Pregler et al. 2015; Geneva et al. 2018).

While numerous studies have found that eDNA can be more sensitive than traditional survey methods when surveying rare species (e.g., Robinson et al. 2019; McColl-Gausden et al. 2020; Nester et al. 2023), we failed to detect bridle shiners with eDNA methods at two sites where we caught fish at low abundances (n = 1 to 3 fish). At one of these sites, OSSIPE, we successfully detected bridle shiners with eDNA the following year. We did not conduct seine and eDNA surveys within the same site visit due to logistical constraints, so it is possible that there were no bridle shiners present at the time of eDNA sampling. We discovered that, while we were able to detect bridle shiner DNA from June to October across all sites, we were only able to detect bridle shiner DNA in mid-June at our long-term sites OSSIPE and PRESUM-01 (Figure 1.3). While we had expected detection probabilities to decrease in the winter due to increased water volume and decreased fish metabolic rate (USFWS 2020), we had also expected to be able to detect bridle shiner DNA throughout the summer at sites where the fish were abundant. It is

possible that bridle shiners were only using these two sites during the spawning period or moving between nearby habitat patches during the summer. Both sites were along river backwaters with variable flow, and it is unclear whether this may have influenced behavior.

Little is known about bridle shiner movement patterns, but other small cyprinids travel both downstream and upstream to nearby habitat patches (Goforth and Foltz 1998; Johnston 2000). Research on bridle shiner movement patterns in Maine would allow us to determine whether eDNA methods failed to detect bridle shiners when they were in fact present or whether these non-detections indicate periodic movements away from surveyed habitat patches. For example, bridle shiners may have been moving between nearby habitat patches in response to seasonal changes in plant cover. Further research is needed to determine the extent of bridle shiner movements between seasons and their overwintering behavior, and to explore potential behavioral differences between populations in pond and stream habitats.

There are advantages and limitations to surveying with eDNA and seine netting, and determining which method to use will depend on multiple factors. The primary advantages of eDNA sampling are that it is non-invasive, is less destructive to habitat than seine netting, and minimally impacts sensitive or imperiled species with small population sizes (Valentini et al. 2016; Nester et al. 2023). eDNA methods will also identify rare species regardless of age or body size (Nester et al. 2023). Seine netting surveys, however, are limited to times of the year when young-of-year fish are more easily identifiable (Jensen and Vokoun 2013). These surveys require more personnel than eDNA surveys and require at least one person who can distinguish bridle shiners from similar species, while eDNA surveys only require personnel who can broadly identify bridle shiner habitat.

Another advantage of using eDNA is that sampling kits can be sent to volunteers or landowners. Other studies have demonstrated that community scientists can collect high-quality eDNA samples, and that these surveys can be conducted over large spatial scales with little temporal variation (e.g., Biggs et al. 2015; Larson et al. 2020). This could be of particular use in Maine, where many smaller waterbodies are surrounded by private property and are only accessible from the shoreline or by canoe or kayak. We found that nearly all landowners were willing to allow us access to waterbodies for eDNA sampling. eDNA samples, however, are sensitive to cross-contamination, with even a slight contamination becoming amplified in later steps of sample processing (Quinn et al. 2018; McColl-Gausden et al. 2020). When eDNA is collected from sites of unknown occupancy status, the prevalence of both false negative and false positive detections (e.g., from cross-contamination) is not known. Occupancy models provide unbiased estimates of occupancy and the probability of false negatives, but only when the probability of false presence is low (Lahoz-Monfort et al. 2016).

We modeled differences in estimated occupancy and detection between survey methods and across all eDNA surveys. We were able to validate six of seven positive bridle shiner eDNA detections with seine net captures but were unable to verify results from the seventh site (GWORKS; Figure 1.3). Over the course of the study, we were able to visually confirm the presence (or likely presence) of bridle shiners at nearly every site where we detected them with eDNA, regardless of whether we surveyed with a seine. We did not see any bridle shiners at GWORKS, and only one qPCR replicate detected bridle shiner DNA at this site. This detection could therefore represent a false positive. However, we also only detected one positive qPCR replicate at site SEBAGO-01 (Figure 1.3) and we successfully verified this detection by capturing bridle shiners. Therefore, we cannot assume that samples with one qPCR detection are

false positives, especially since eDNA concentrations (and therefore detection probability) are expected to differ between sites (Lahoz-Monfort et al. 2016). Additionally, Ficetola et al. (2015) found that the practice of excluding such ambiguous samples from occupancy analyses can introduce significant bias to estimates of occupancy and detection probability.

The GWORKS sample may have represented a false positive in another sense in that it could have contained bridle shiner DNA that had been transported downstream (Roussel et al. 2015). While eDNA in still-water systems does not disperse far from its source (Eichmiller et al. 2014; Goldberg et al. 2018), eDNA in streams and rivers can travel multiple kilometers downstream. Deiner and Altermatt (2014), for example, detected target DNA approximately 20-km downstream of where the organisms were present. Such detections can be problematic when determining habitat use, as the habitat characteristics at the eDNA sampling location will not necessarily reflect the habitat the organism is using.

There is no universal methodology for eDNA sampling because experimental designs and protocols must be adjusted to accommodate different study systems and target species (Takahashi et al. 2023). We chose a 0.7-µm filter pore size to survey bridle shiners because we needed to balance higher eDNA retention at smaller pore sizes (Eichmiller et al. 2014) with the increased time needed to filter larger sample volumes. While smaller filter pore sizes (e.g., 0.2µm) capture a wider size range of eDNA particles, only low volumes of water can be passed through each filter (Turner et al. 2014; Goldberg et al. 2018). We have shown that collecting four 1-L replicates (or 4-L total) of water per habitat patch will result in a greater than 95% probability of capturing bridle shiner DNA (given that bridle shiners are present) when using this filter type. However, collecting this volume of water, plus a 1-L control, from each site and transporting it to a filtering location can be challenging.

We were able to successfully detect bridle shiner eDNA with pore sizes as large as 2.7-µm, so future studies may consider conducting hierarchical occupancy modeling on samples filtered with larger pore sizes. We saw no benefit from using finer pore sizes, and increasing filter pore size could allow for filtration of larger volumes of water in the field (Turner et al. 2014; Hinlo et al. 2017; Takahashi et al. 2023), which would save processing time. While filtering water at a site can increase the risk of sample contamination, it also yields higher DNA copy numbers than filtering after short or long-term sample storage (Hinlo et al. 2017).

Collecting multiple eDNA samples is less time-consuming than a seine survey. Using eDNA can enable more site visits per sampling day and accurate monitoring across broad spatial scales (Turner et al. 2014; Valentini et al. 2016; Deiner et al. 2016). Overall, however, we found that the combination of eDNA collection and processing (e.g., filtering samples and decontaminating sampling equipment) was more time consuming than seine netting. Increasing filter pore size and filtering in the field would eliminate the need for much of the sampling equipment we used in this study and would save processing time. Combining sample replicates taken from multiple locations across a site would also reduce the need for separate sampling kits and would decrease filtration time (Goldberg et al. 2018).

Streamlining the eDNA survey protocol for bridle shiners will be highly beneficial, as these surveys can be conducted with fewer personnel than seine net or electrofishing surveys. Seine surveys can also result in mortality of both bridle shiners and other small-bodied or young-of-year fish. Seine netting does, however, allow managers to monitor bridle shiner health and abundance at a site in real time (Nester et al. 2023). Determining relative abundance from eDNA is sometimes possible, but factors other than abundance affect the concentration of DNA in a sample. While studies have shown positive correlations between eDNA concentration and the

relative abundance of a species (e.g., Thomsen et al. 2012; Lacoursière-Roussel et al. 2016a; Rourke et al. 2021), differences in eDNA transport, degradation, and production between study systems and organisms preclude us from reliably estimating abundance in all instances (Rourke et al. 2021; Wood et al. 2021). For example, eDNA concentration within a water body is determined by fish distribution as well as abundance and is not spatially homogeneous (Takahara et al. 2012; Eichmiller et al. 2014). Sites at which we detected bridle shiner DNA across multiple replicates may have supported higher abundances than sites where we only detected DNA in one replicate. Alternatively, differences in DNA concentration and detection could have resulted from sampling high-use areas of a habitat patch at one site and low-use areas at another (Eichmiller et al. 2014).

This study provided a baseline for bridle shiner abundance at sites where we seine netted. Lake surveys conducted by G.P. Cooper (1939), which form the basis for most of our historic bridle shiner knowledge, only described relative abundance within a waterbody ("abundant," "common", "rare", or absent). Few reports contained the number of bridle shiners caught, the precise location where they were found, and a quantification of survey effort. It is not known why bridle shiners are persisting in some of their historically occupied sites but not in others. Sites such as Stanley Pond (STANPD-01/02/03), Trafton Pond (TRAFPD-01/02), and Spectacle Pond (SPECPD) have abundant bridle shiner habitat, but we could not find evidence of current occupancy. Other sites, such as the Jordan River outlet (JORDAN) and Josie's Brook (JOSIES), had been invaded by variable-leaved watermilfoil (*Myriophyllum heterophyllum*). Comparing habitat variables between historic bridle shiner sites and currently occupied locations could reveal shifts in habitat suitability over time and could allow us to predict other areas where bridle shiner populations are likely to persist.

1.4.1. Conclusions

Collectively, our results provide evidence that bridle shiner populations are declining in Maine. Despite rediscovering bridle shiners at 11 of their historically occupied sites and documenting bridle shiners in four new waterbodies, we were unable to detect bridle shiners at 21 of 32 sites known to have once supported the species. These results suggest a loss of approximately 65% of known Maine bridle shiner populations. We determined that both eDNA and seine net surveys are viable options for monitoring bridle shiners in Maine, and that the eDNA methods used in this study can be further streamlined to reduce the time and cost of future surveys.

CHAPTER 2

USING LOCAL AND LANDSCAPE-SCALE HABITAT VARIABLES TO PREDICT BRIDLE SHINER (*NOTROPIS BIFRENATUS*) DISTRIBUTION IN MAINE AND NEW HAMPSHIRE

2.1. Introduction

Freshwater ecosystems are among the most critically threatened habitats on Earth, and within these ecosystems, freshwater fish are among the most imperiled animals (Dudgeon et al. 2006; Jelks et al. 2008). As of 2008, nearly 40% of North America's freshwater and diadromous fish species were considered vulnerable, threatened, or endangered (Jelks et al. 2008). Anthropogenic environmental impacts such as habitat loss, degradation, and fragmentation, have increased in prevalence over recent decades (Paul and Meyer 2001; Walsh et al. 2005). Small-bodied fish species such as minnows and darters are especially vulnerable to the impacts of habitat alteration and destruction (Whittier et al. 1997; Olden et al. 2007). Changes to natural hydrologic regimes, increased turbidity and pollution, and the introduction of invasive species frequently pose threats to small-bodied freshwater fishes (Angermeier 1995; Bunn and Arthington 2002; Jelks et al. 2008; Gray et al. 2016).

Recent declines and extirpations of native minnows in the northeastern United States and Canada have been linked to both habitat loss or degradation and the stocking of non-native predatory sport fishes (Whittier et al. 1997). For example, Whittier et al. (1997) found that the introduction of largemouth bass (*Micropterus salmoides*), smallmouth bass (*M. dolomieu*), and northern pike (*Esox lucius*) was the most consistent factor related to declines in native minnow species richness in northeastern lakes. Increased human activity within a watershed has also been related to decreased minnow species richness. The development of lake shorelines and stream

floodplains alters both the physical habitat and nutrient cycling (Scheuerell and Schindler 2004). This is due to practices such as clearing woody debris and vegetation from water bodies (Whittier et al. 1997). Anthropogenic effects are further compounded as catchment basins integrate threats and disturbance from their surrounding landscapes (Dudgeon et al. 2006; Olden et al. 2007).

Small-bodied fish are especially threatened by extirpation when they occupy limited geographic ranges, have few occurrences, or have a highly fragmented distribution (Fagan 2002). These specialist species may already have lower populations and be especially vulnerable to disturbance and environmental stochasticity (Angermeier 1995). Shifts in water chemistry, temperature, and vegetation resulting from anthropogenic impacts can all contribute to local extirpations of specialist species (Angermeier 1995).

Protecting a species at the peripheral edge of its range can pose additional challenges, as populations may be more isolated than in the core of their range (Haak et al. 2010; Lamothe and Drake 2020). Populations at the northern edge of their range may also utilize different habitats than populations further south (Haak et al. 2010; Lamothe and Drake 2020). Northernmost peripheral populations may have a heightened importance to the persistence of a species as climate change shifts biomes poleward (Gibson et al. 2009). Spatially isolated peripheral populations of species with poor dispersal ability and short generation time have a higher likelihood of genetic divergence and differentiation (Lesica and Allendorf 1995).

High-latitude states, such as Maine and New Hampshire, and the southern Canadian provinces support many temperate fish species at the northernmost limit of their range and boreal species at the southern limit of their range. Several of these species are locally at risk despite being common in other parts of their range (Gibson et al. 2009). Monitoring these peripheral

populations is critical to their species' conservation as rising temperatures shift plant communities northward. Finding peripheral populations of rare or declining species, especially small-bodied and/or cryptic species, is challenging, and detection can be extremely low when sampling over large areas (Guisan et al. 2006).

One strategy to develop targeted surveys for rare species is species distribution modeling (SDM). SDM statistically associates species occurrence data with environmental variables in order to evaluate habitat suitability (Riaz et al. 2020). SDMs can be used to discover new populations of rare species by identifying areas with suitable habitat (Riaz et al. 2020). Spatially explicit habitat models also result in greater survey efficiency when compared with simple or stratified random sampling over large areas (Guisan et al. 2006). Many modeling strategies have been implemented to predict species occurrences, including generalized linear models ("GLMs"; e.g., Carlos-Júnior et al. 2020), maximum entropy modeling ("Maxent"; Phillips et al. 2006, Elith et al. 2011), and random forests ("RF"; Breiman 2001, Hengl et al. 2018, Valavi et al. 2021). Maxent is one of the most widely-used SDM techniques because it can be used to model presence-only data (Phillips et al. 2006), is robust to the small sample sizes typical of rare species surveys (Kaky et al. 2020), and can model non-linear relationships between species presence and predictor variables (Elith et al. 2011). The predictions of individual models can be highly variable, and the choice of modeling method is known to impact model outcomes and accuracy (Araújo and New 2007). Ensemble modeling, or combining model predictions, can amplify the patterns found across models and produces more robust predictions (Marmion et al. 2009).

The bridle shiner is a small-bodied, specialist minnow native to the eastern United States and Canada. Bridle shiners depend on clear, shallow water with abundant aquatic vegetation, and

are highly sensitive to the changes in water quality, turbidity, and plant cover that result from anthropogenic disturbance (Cooper 1985; Gray et al. 2016). Their distribution among reaches within a watershed is naturally patchy due to their specific habitat requirements. Because natural movement within bridle shiner metapopulations is already limited, additional anthropogenic barriers to movement can further isolate subpopulations and increase the risk of local extirpation (Johnston 2000). Historically, Maine's Saco River watershed marked the eastern limit of the bridle shiner's range.

The bridle shiner is thought to be declining dramatically throughout most of its native range, and probably has been extirpated entirely from the state of Maryland (Kilian et al. 2011). Bridle shiners were once abundant in Delaware, Maryland, New Jersey, and Pennsylvania, but populations have declined as urbanization and industrial and agricultural development have increased (Cooper 1985). There are few or no known bridle shiner populations left in Virginia, North Carolina, and South Carolina (Geneva et al. 2018). This species now receives legal protection or concern status in thirteen states and two provinces (COSEWIC 2013; Hammerson 2021). Bridle shiners are listed as state Threatened in New Hampshire (New Hampshire Fish and Game Dept. [NHFGD] 2015) and as a Species of Special Concern in Maine (Maine Dept. of Inland Fisheries and Wildlife [MDIFW] 2021). They are considered a Species of Greatest Conservation Need in both states (MDIFW 2015, NHFGD 2015).

The two objectives of this study were to 1) assess small-scale bridle shiner habitat selection within a waterbody, and to 2) inform bridle shiner conservation at the regional scale by modeling their distribution across southern Maine and New Hampshire. Recent bridle shiner surveys in Maine (2021-2022; Chapter 1) present a unique opportunity to assess habitat selection at both the local and regional scale. Combining the Maine surveys with a presence-absence

dataset from New Hampshire allowed us to look at patterns of bridle shiner occupancy over time in the northeastern-most part of their range. We used an ensemble SDM approach to characterize both the current (2000-2022) and historic (1898-1999) ranges of the bridle shiner (*Notropis bifrenatus*) within these two states.

2.2. Methods

2.2.1. Study Area

The historic bridle shiner range in Maine and New Hampshire falls within two Level III Ecoregions: the Northeastern Coastal Zone and the Northern Appalachian and Atlantic Maritime Highlands (Wiken et al. 2011). Much of this area was formerly glaciated, and most of the lakes were formed by glaciers (Wiken et al. 2011; Deeds et al. 2020). The Northern Appalachian region is dominated by mixed hardwood and spruce-fir forests and is transitional between the northern boreal forests and the deciduous forests of New England (Wiken et al. 2011). Waterbodies along the coast are impacted by marine-derived sediments known as the Presumpscot Formation, and coastal Maine contains much of the state's agriculture and human population because of this (Deeds et al. 2020). This is significant because agricultural and developed land use lead to increased erosion, excess nutrient loading, and an influx of road salt, which can all be significant stressors on lake ecosystem health (Soranno et al. 2015; Sutherland et al. 2018; Deeds et al. 2020). The Coastal region defined by Deeds et al. (2020) has the warmest average temperatures in Maine, and therefore has the shortest period of winter ice-over and the longest period of summer bioproductivity.

2.2.2. Maine Bridle Shiner Surveys

We surveyed Maine bridle shiner populations using seine netting and environmental DNA (eDNA) over the summer and fall of 2021 and 2022. Environmental DNA and seine

netting protocols are described in Chapter 1. Prior to collecting water, we surveyed aerial imagery from locations where bridle shiners had been reported between 1937 and 2010. We chose water collection subsites based on qualitative habitat suitability (Jensen and Vokoun 2013; Pregler et al. 2015, 2019) and sampled 41 subsites within 30 sites between 1 June and 16 July 2021. We then seined 29 sites that produced a historical (1930s-1940s) or recent (1990s-2010s) record of bridle shiner occurrence. In 2022, we created a preliminary habitat suitability model based on bridle shiner habitat preferences in published literature (Chapter 1) and surveyed 58 locations in 46 waterbodies.

We detected bridle shiners at 17 locations out of the 97 locations surveyed in Maine. We used the 80 locations where we failed to detect bridle shiners as absences in SDMs of the current bridle shiner distribution only, as we could not be certain that bridle shiners had never occupied those areas. We removed records (n = 4) from locations where MDIFW biologists suspected bridle shiners had been introduced (Marshall Brook in Acadia National Park; Doering et al. 1995) or had been misidentified (i.e., areas well outside of the known bridle shiner range but within the range of the visually similar blacknose shiner, *N. heterolepis*).

2.2.3. New Hampshire Bridle Shiner Surveys

NHFGD conducted fisheries surveys between 2005 and 2022 as part of eight different projects (NHFGD 2015). Capture methods included seine netting, boat electrofishing, backpack electrofishing, dip netting, and minnow trapping (M. Carpenter, personal observation). Surveys conducted for other fish, such as brook trout, provided presence-only bridle shiner data, while surveys conducted specifically for bridle shiners also noted sites where bridle shiners were absent or extirpated. Surveys conducted specifically for bridle shiner mostly used dip netting, seine netting, and minnow traps.

We aggregated New Hampshire records by year in ArcGIS Pro (version 3.1.2, Environmental Systems Research Institute, Redlands, CA) so that locations with repeated surveys were only represented by one point (n = 147 locations). NHFGD recently introduced bridle shiners into a small pond, so we considered this site the only known historic absence and did not include it in the current population models. There were eight lakes or ponds where bridle shiners were reported as extirpated in New Hampshire's Wildlife Action Plan (NHFGD 2015): we included these as historic presence locations in the SDMs and used the lake centroid coordinates.

2.2.4. Local Habitat Variables

2.2.4.1. Habitat Data Collection. We measured habitat characteristics at 98 sites in Maine in 2021 and 2022. At sites where we seine netted (2021), we recorded water depth in three locations for each seine sample and measured total dissolved solids (ppm), water temperature (°C), and conductivity at each sampled habitat patch. We also determined the sediment type(s) and dominant plant species for each seine sample and took photos of the site and the deployed seine net for reference. We collected samples or recorded the name of all submerged, emergent, and floating plant species at each location. We made note of plants that we did not collect (such as large water lilies [*Nymphaea* spp. and *Nuphar* spp.] and pickerelweed [*Pontederia cordata*]), and stored samples in a freezer until they could be thawed and identified.

In 2022, we recorded habitat information at each location where we collected eDNA samples. We visually estimated the proportion of submerged, floating, and emergent vegetation and the proportion of open water at each site. We then estimated the proportion of total submerged vegetation made up of simple-leaved, complex-leaved, and mat-forming/grass-like plants, and the proportion of total emergent vegetation composed of persistent vegetation (i.e.,

grasses, rushes, and sedges), broad-leaved deciduous vegetation, and cattails (Nohner and Diana 2015). We then visually estimated the proportion of the site composed of organic substrate, small (inorganic) substrates less than 2-mm in diameter, and large inorganic substrates greater than 2-mm in diameter (Lamothe and Drake 2020). We measured total dissolved solids (ppm), water temperature (°C), and conductivity (µS/cm) at each site. We did not directly measure the proportion of organic, small inorganic, and large inorganic substrates or the proportion of vegetation types during sampling in 2021, so we used photographs of the sites to estimate these values.

2.2.4.2. Local Habitat Modeling. We used binomial generalized linear models (GLMs) to identify local-scale environmental variables associated with bridle shiner presence in Maine. We collected or calculated 35 environmental variables from each of the 95 sites where we collected eDNA and/or seined in 2021 and 2022 (Table 2.1). We counted the number of dams within a 2-km radius of a sampling location (within the same drainage; Pregler et al. 2019), and used the 2016 National Land Cover Database (NLCD) Tree Canopy Cover dataset to determine the percent canopy cover at each sampling location (Coulston et al. 2012). We also calculated the proportion of seven 2019 NLCD land cover classes (agricultural, developed, total forest, mixed forest, deciduous forest, coniferous forest, and wetland/open water) within each site's HUC12 sub-watershed (Table 2.1; Dewitz and U.S. Geological Survey [USGS] 2021, USGS 2021). In addition to using the HUC12 delineations from the USGS Watershed Boundary Dataset (USGS 2021), we used a digital elevation model (DEM; USGS 1998) and ArcGIS Pro's Hydrology toolset to calculate the upstream drainage area of each sampling location. We hypothesized that land use within each site's upstream drainage area would have more of an influence on water quality than land use in its surrounding HUC12, which also included the area downstream of a

site. We used the Index of Ecological Integrity (IEI) developed by the North Atlantic Landscape Conservation Cooperative (McGarigal et al. 2018) as a proxy for site disturbance. We included the measurements of conductivity (µS/cm) and total dissolved solids (ppm) that we had collected in the field along with our estimates of site substrate and plant cover described above (Table 2.1).

Table 2.1 Covariates used to determine local habitat effects on bridle shiner presence in Maine.

Category	Variable	Description	Source
Indices of disturbance	dams	Number of dams within 2 km of sampling location	State dam point locations (MAODS and MADCR 2012; ME DEP 2022;
	IEI	Index of Ecological Integrity	North Atlantic Landscape Conservation Cooperative (McGarigal et al. 2018)
Land use: HUC12	HUC12.area	Area in sq. km of each HUC12 unit	Watershed Boundary Dataset (USGS)
	Prop.ag.HUC12	Proportion of agricultural area within each site's HUC12	Derived from National Land Cover Database (Dewitz and USGS 2021) and Watershed Boundary Dataset (USGS)
	Prop.cfor.HUC12	Proportion of coniferous forest within each site's HUC12	
	Prop.devel.HUC12	Proportion of developed area within each site's HUC12	
	Prop.dfor.HUC12	Proportion of deciduous forest within each site's HUC12	
	Prop.for.HUC12	Proportion of total forest within each site's HUC12	
	Prop.mfor.HUC12	Proportion of mixed forest (including forested wetland) within each site's HUC12	
	Prop.wetl.HUC12	Proportion of freshwater wetland and open water within each site's HUC12	

Table 2.1 Continued.

Land use: upstream	Drainage.Area	Drainage area for each location in square km calculated using the Watershed tool in ArcGIS Pro	Derived from DEM (USGS 1998)
	Prop.Ag	Proportion of agricultural area within each drainage	Derived from NLCD (Dewitz and USGS 2021)
	Prop.Cforest	Proportion of coniferous forest within each drainage	
	Prop.Devel	Proportion of developed area within each drainage	
	Prop.Dforest	Proportion of deciduous forest within each drainage	
	Prop.Forest	Proportion of total forest within each drainage	
	Prop.Mforest	Proportion of mixed forest (including forested wetland) within each drainage	
	Prop.Wtl	Proportion of freshwater wetland and open water within each drainage	
Plant types & cover	Canopy	Percent canopy cover	NLCD 2016 Tree Canopy Cover (Coulston et al. 2012)
	prop.Eveg	Proportion of site dominated by emergent aquatic vegetation	Estimated at site (2022) or from site photos (2021)
	prop.site.Eveg.broad	Proportion of site dominated by broad-leaved deciduous emergent aquatic vegetation	
	prop.site.Eveg.cat	Proportion of site dominated by emergent aquatic vegetation (cattails)	
	prop.site.Eveg1	Proportion of site dominated by persistent emergent aquatic vegetation	
	prop.Fveg	Proportion of site dominated by floating aquatic vegetation	

Table 2.1 Continued.

Plant types & cover	prop.Sveg	Proportion of site dominated by submerged aquatic vegetation	Estimated at site (2022) or from site photos (2021)
	prop.site.Sveg.complex	Proportion of site dominated by complex submerged aquatic vegetation	
	prop.site.Sveg.grass	Proportion of site dominated by mat-forming or grass-like submerged aquatic vegetation	
	prop.site.Sveg.simple	Proportion of site dominated by simple submerged aquatic vegetation	
	prop.open.water	Proportion of site dominated by open water (no vegetation)	
Substrate	prop.large.sub	Proportion of large substrates	
	prop.org.sub	Proportion of organic substrate	
	prop.sm.sub	Proportion of small substrates	
Water quality	conductivity	Conductivity (µS/cm)	Measured at site
	TDS	Total dissolved solids (ppm)	
Waterbody type	WBType	LakePond or StreamRiver	Categorized by sampling location

We fit GLMs to site data in Program R (version 4.3.0; R Core Team 2023). We first scaled each numeric variable about its mean and standard deviation and generated ten random seed numbers to use in *k*-fold cross-validation. Each seed allowed the *dismo* package (version 1.3-9; Hijmans et al. 2022) to randomly partition the data into five folds using the *kfold* function. This ensured that 80% of the data would be used for training and that 20% would be used for model testing, and that data points would be randomly assigned to the testing and training groups in ten different ways. We then ran the 10-fold cross-validated GLM and generated a correlogram

of the Pearson correlation between numeric variables (Figure 2.1). We ranked variables by their average variable importance across folds using the R package *vip* (version 0.3.2; Greenwell and Boehmke 2020), and eliminated low-ranking variables that were highly correlated with higher-ranking variables ($|r| \geq 0.70$; Dormann et al. 2013). We then simplified the model by running a genetic algorithm ten times in package *glmulti* (version 1.0.8; Calcagno 2020). *Glmulti* uses Akaike's Information Criterion (AIC; Akaike 1974) to determine which linear combination of variables results in the best model fit. We fit a final GLM with only the reduced variables from the top *glmulti* model and calculated the model's area under the receiver operating characteristic curve (AUC; Fielding and Bell 1997). Finally, we ran a χ^2 analysis of deviance to determine which variables had the greatest influence on model fit.

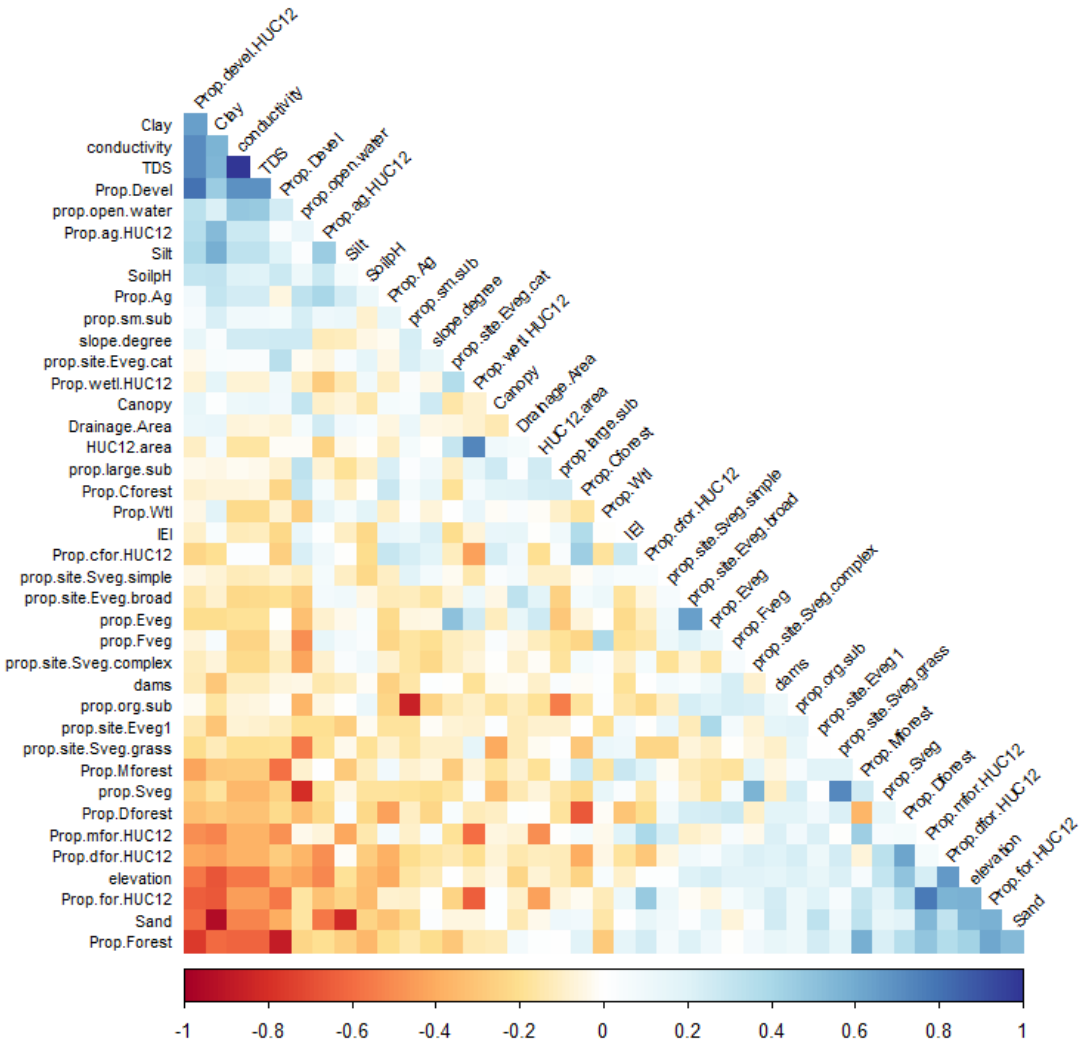


Figure 2.1 Correlogram of the Pearson correlations (*r*) between 34 continuous local habitat variables. Darker reds denote strong negative correlations and darker blues denote strong positive correlations.

2.2.5. Species Distribution Models

2.2.5.1. Raster and Presence-Absence Data Preparation. We selected 24 environmental variables available in GIS repositories to include in our SDM models (Appendix E). We limited the extent of the models by buffering the known historic range of bridle shiners in New

Hampshire and Maine (i.e., the Saco and Merrimack HUC6 basins) by 50-km to account for the likely non-detection of fish at the edges of their range (Sutton et al. 2015). We created one raster at this extent for each variable, and projected all rasters to NAD83 / UTM Zone 19N using either ArcGIS Pro or the *gdalUtilities* package in R (soil data only; version 1.2.4 O'Brien 2023). We then resampled each raster by using the 2019 NLCD as a snap raster. We also applied the NLCD cell size (approximately 30-m x 30-m) to the rasters. We used all 24 covariates to determine which land cover classes, landscape position variables, and substrate variables to retain in the final SDMs. We scaled all rasters of continuous variables by subtracting the raster's mean value from each cell and then dividing by the standard deviation.

We used the most recent survey at a site to determine which sites were occupied during the historic (1898-1999) and current (2000-2022) time periods (Appendix F). If bridle shiners were documented at a site after the year 2000, we presumed that they were also historically present at the site. We considered a site to be currently occupied if bridle shiners were found there during the most recent site survey between 2000 and 2022. There were 17 lakes and ponds with historic bridle shiner presence but no survey coordinates: we used the center point of these in the SDMs unless the location of a suitable habitat patch was known (n = 7 lakes/ponds).

2.2.5.2. Landscape-Scale Variables. We included the landscape position variables catchment position (*catchment*), distance from the coast (km; *coast*), elevation (m; *elev*), marine limit (*marine*), and slope (degrees; *slope*). We determined the catchment position of a site following the method of Pregler et al. (2019) and using data from the Northeast Aquatic Habitat Classification System (NAHCS; Olivero and Anderson 2008) to assign Strahler stream order (Strahler 1957). We derived site slope and elevation from DEMs and calculated the distance from the coastline in kilometers. We considered areas under 128-m in elevation along the

coastline to be within the marine limit, and therefore influenced by marine-derived sediments (Deeds et al. 2020). We hypothesized that bridle shiners would be positively associated with water bodies closer to headwaters (catchment positions 1a, 1b, and 2; Olivero and Anderson 2008, Pregler et al. 2019), in areas of low elevation and with lower slopes, in areas closer to the coast, and within the marine limit.

We divided the model extent into 177 hexagonal cells (circumcircle radius = 10-km, area = 259.8-km^2), then used the 2020 Biophysical Settings dataset from the Landscape Fire and Resource Management Planning Tools (LANDFIRE; Rollins 2009, Blankenship et al. 2021) program to calculate the proportion of fourteen forest types within each hexagon (Table 2.2). This allowed us to explore variations in forest type along north-south, east-west, and coast-mountain gradients.

Finally, we included five substrate variables in our models. We obtained four of these layers (clay content, sand content, silt content, and soil pH) from the International Soil Reference and Information Centre's (ISRIC) Soil Data Hub (Poggio et al. 2021). We included the clay, sand, and silt content of the uppermost 5-cm of soil (in g/kg; Poggio et al. 2021) to locate water bodies with sand or silt substrates (Adams and Hankinson 1928). We also included lithology classes (fine glacial lake sediment, coarse glacial outwash, and fine coastal sediment and alluvium) as these are the parent materials of soil substrates and remain stable over long time scales (Theobald et al. 2015). Finally, we included soil pH as a proxy for waterbody pH as many cyprinid species are intolerant of highly acidic water (Laerm et al. 1980; Rahel and Magnuson 1983), and acidity may limit the distribution of some species (Laerm et al. 1980). Cell values for soils below water bodies were not included in the original ISRIC soil rasters, so we interpolated

these values within a 200-m shoreline buffer of each waterbody using the Empirical Bayesian Kriging tool in ArcGIS Pro (default model and search neighborhood parameters).

Table 2.2 Rasters of landscape-scale variables used to model bridle shiner distribution in Maine and New Hampshire.

Category	Variable	Description	Source
Landscape position variables	catchment*	Catchment position: 6 stream/river size classes	Northeast Aquatic Habitat Classification System (Olivero and Anderson 2008)
	coast	Distance from coast (km)	Derived from National Wetlands Inventory (USFWS 2022)
	elev	Elevation (m)	Digital Elevation Model (USGS 1998)
	marine†	Marine limit	Derived from Digital Elevation Model (USGS 1998)
	slope	Slope (°)	
Land cover variables	for.1920	Proportion of Laurentian-Acadian Northern Hardwoods Forest in cell*	LANDFIRE Biophysical Settings (Rollins 2009; Blankenship et al. 2021)
	for.1921	Proportion of Northeastern Interior Dry-Mesic Oak Forest in cell	
	for.1922	Proportion of Northern Atlantic Coastal Plain Hardwood Forest in cell	
	for.1924	Proportion of Laurentian-Acadian Northern Pine(-Oak) Forest in cell	
	for.1925	Proportion of Laurentian-Acadian Pine-Hemlock-Hardwood Forest in cell	
	for.1926	Proportion of Central Appalachian Dry Oak-Pine Forest in cell	
	for.1927	Proportion of Appalachian (Hemlock-)Northern Hardwood Forest in cell	
	for.1928	Proportion of Acadian Low-Elevation Spruce-Fir-Hardwood Forest in cell	
	for.1929	Proportion of Acadian-Appalachian Montane Spruce-Fir Forest in cell	

Table 2.2 Continued.

Land cover variables	for.1930	Proportion of Central Appalachian Pine-Oak Rocky Woodland in cell	LANDFIRE Biophysical Settings (Rollins 2009; Blankenship et al. 2021)
	for.1931	Proportion of Northern Atlantic Coastal Plain Maritime Forest in cell	
	for.1941	Proportion of North-Central Interior Wet Flatwoods in cell	
	for.1980	Proportion of Boreal Jack Pine-Black Spruce Forest in cell	
	for.1981	Proportion of Northeastern Interior Pine Barrens in cell	
Substrate variables	clay	Clay content of soil (g/kg) at 0-5cm	SoilGrids (Poggio et al. 2021)
	pH	Soil pH (pH * 10) at 0-5cm	
	sand	Sand content of soil (g/kg) at 0-5cm	
	silt	Silt content of soil (g/kg) at 0-5cm	
	lith‡	Lithology: 4 classes + water	(Theobald et al. 2015)

Note: * NAHCS stream orders: 1a:Headwater: 0<3.861 sq.mi, 1b:Creek: >=3.861<38.61 sq.mi., 2:Small River: >= 38.61<200 sq.mi., 3a:Medium Tributary River: >=200<1000 sq.mi., 3b:Medium Mainstem River: >=1000<3861 sq.mi., 4:Large River: >=3861<9653 sq.mi., 5:Great River: >=9653 sq.mi.
† Categorical: Falling within (1) or outside of (0) the marine limit as defined by Deeds et al. (2020)
‡ 5 classes of lithology: water (999), glacial till coarse (11), glacial lake sediment fine (13), glacial outwash coarse (14), alluvium and coastal sediment fine (19)

2.2.5.3. Exploratory Machine-Learning Models (Maxent and Random Forest). We used maximum entropy (Maxent; Phillips et al. 2006) and random forest models (Breiman 2001) to explore the relationships between the 24 raster habitat covariates and current and historic bridle shiner distributions in Maine and New Hampshire. These models are both machine learning methods that can fit complex nonlinear relationships (Breiman 2001). We stacked the 24 rasters using package *terra* (version 1.7-28; Hijmans 2023) and loaded the presence-absence data into package *SDMtune* (version 1.2.1; Vignali et al. 2020). We followed the *SDMtune* stepwise variable selection and model-tuning protocol described by Vignali et al. (2020) to find the most

parsimonious group of predictor variables for each type of model and for historic and current survey data.

SDMs compare habitat variables at presence locations with the available habitat at randomly generated background points over the model's spatial extent. We added known absences from each time period (n = 1 historic, n = 116 current absences) to randomly-generated background points (n = 10,000 total points; Barbet-Massin et al. 2012) using *SDMtune*. We weighted presence and pseudo-absence points equally (Barbet-Massin et al. 2012). We then split presence and absence-background points into training (60%), validation (20%), and testing (20%) datasets for model cross-validation (*sensu* Vignali et al. 2020). We calculated each model's area under the receiver operating characteristic curve (AUC; Fielding and Bell 1997) after each tuning step using the validation dataset, and then used the held-apart testing dataset to calculate the final model AUC (Vignali et al. 2020). We used the default permutation value of 10 for all analyses.

We first ran default models using the training dataset with no cross-validation. Then, we performed k-fold cross-validation (k = 10 folds; Sutton et al. 2015) using the default model settings. We then plotted the Pearson correlation between all continuous covariates to gauge whether any of the predictor variables were highly correlated ($|r| \geq 0.70$; Figure 2.2; Dormann et al. 2013). We used a data-driven approach to select the predictor variables with the highest explanatory value that were not highly correlated with other variables (Vignali et al. 2020). The *varSel* function in *SDMtune* ranks predictor variables by permutation importance, then performs a leave-one-out Jackknife test to determine which variable within each group of correlated variables will reduce model performance the least when removed. The function iterates through

all the predictor variables until all the remaining correlations fall under the threshold value of 0.70 (Dormann et al. 2013; Vignali et al. 2020).

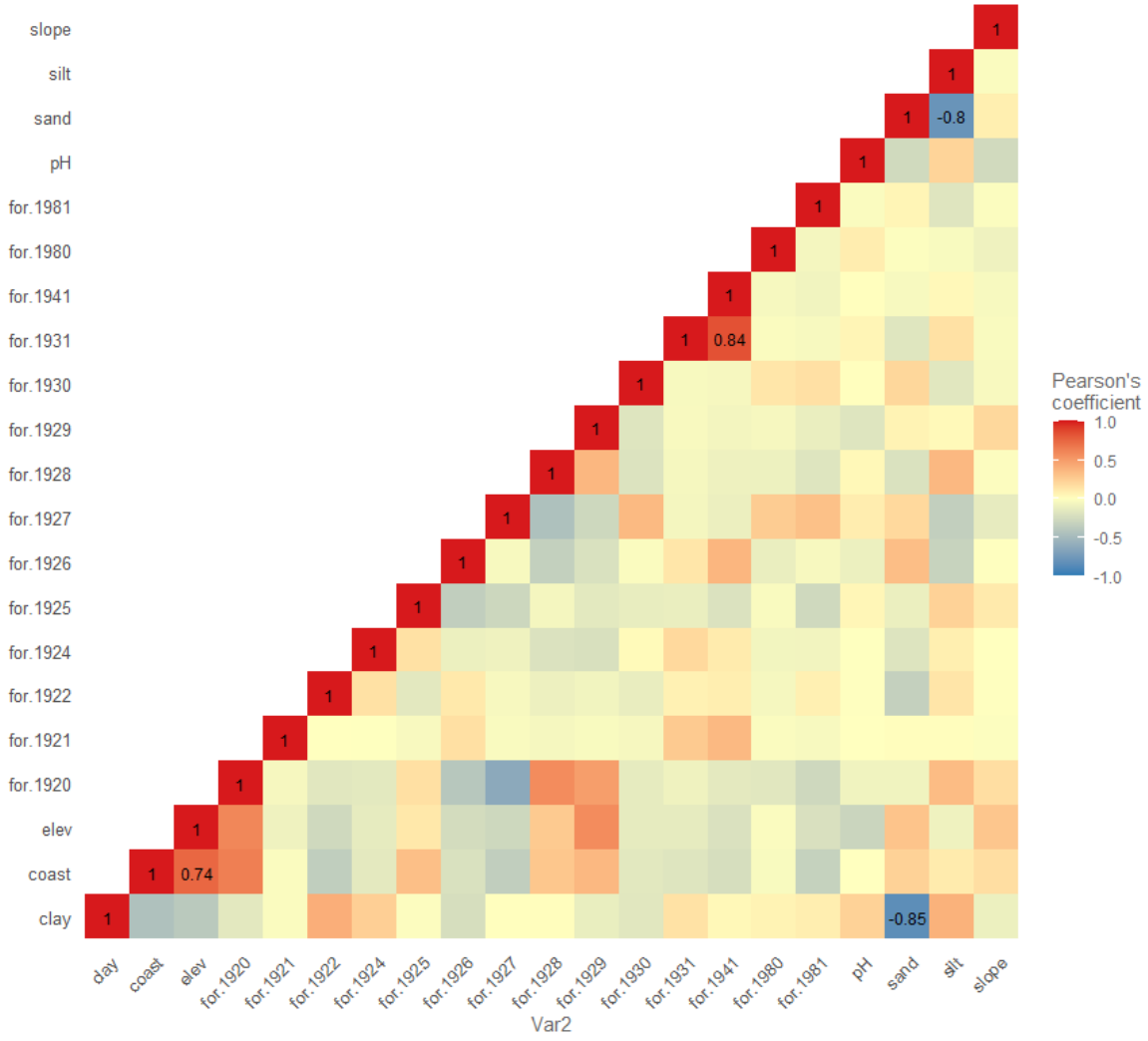


Figure 2.2 Correlogram of the Pearson correlations (*r*) between 21 continuous landscape habitat variables. Pearson coefficients greater than 0.70 are labeled.

Machine learning models have fixed settings, or hyperparameters, that must be defined prior to model training. Optimal hyperparameter values are specific to each dataset, and tuning a machine learning model requires testing multiple configurations of these parameters (Vignali et al. 2020). We tuned Maxent and RF hyperparameters using the *optimizeModel* function in *SDMtune* with default arguments (Vignali et al. 2020). Tunable hyperparameters differ between

model types: typical random forest hyperparameters include the number of trees (*ntree*), the number of candidate features to select at intermediate nodes of a tree (*mtry*), and the minimum number of observations informing each terminal node (*nodesize*; Han et al. 2020). Tunable Maxent hyperparameters include feature class combinations (*fc*; linear, quadratic, product, hinge, and threshold), the regularization multiplier (*reg*), and the number of model iterations (*iter*). The *optimizeModel* function applies a genetic algorithm to optimize the combination of possible hyperparameter values rather than calculating all possible combinations (Vignali et al. 2020). We provided the function with the following list of possible hyperparameter values: *mtry* = 1-10, *ntree* = 500, 700, 900, 1100, 1300, 1500, 1700, or 1900, and *nodesize* = 1-15 for RF models and *fc* = "l", "lq", "lh", "lqp", "lqph", or "lqpht", *iter* = 300, 500, 700, 900, or 1100, and *reg* = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, or 1.0 for Maxent models.

We used the tuned hyperparameters to further optimize model parsimony. The *SDMtune* function *reduceVar* removes predictor variables with low permutation importance (Vignali et al. 2020). This function works similarly to the *varSel* function in that it performs a leave-one-out Jackknife test on each variable and removes variables that fall below the threshold permutation importance value (provided that the removal of the variable does not decrease model AUC; Vignali et al. 2020). We chose a conservative threshold permutation importance value of 1% , then merged the training (60% of data) and validation (20% of data) data subsets to make a larger training dataset with which to test the final exploratory model. We evaluated this model using the held-apart testing dataset (20% of data), which had not been used in earlier steps to tune the model (Vignali et al. 2020).

2.2.5.4. Exploratory Generalized Linear Models. We fit binomial GLMs for both current and historic bridle shiner data. As with the RF and Maxent models, we used current presence

locations (n = 122), current absence locations (n = 116), and a subset of the background pseudo-absence locations generated in *SDMtune* (n = 9884) to model the current bridle shiner distribution. For the historic dataset, we used presence locations (n = 156), absence locations (n = 1), and a larger subset of the pseudo-absence locations (n = 9999). As with the local habitat GLMs, we began by generating ten random seed numbers to use in k-fold cross-validation. We randomly assigned 80% of the points to the training dataset and 20% to the testing dataset using the ten seeds. We did not include a validation subset of the data as with *SDMtune* as there were no hyperparameters to tune.

For each dataset, we then conducted a 10-fold cross-validated GLM with all 24 scaled predictor variables. We then calculated the variable importance to each model using *vip*. We used the average effect size (*Z*-value) to remove less-influential variables that were highly correlated ($|r| \geq 0.70$; Dormann et al. 2013) with highly-influential variables, as we had done with the random forest and Maxent models in *SDMtune*. We then dropped the correlated layers from the raster stack and ran a new GLM with the reduced raster dataset. To approximate the *reduceVar* step used with the Maxent and RF models, we used *glmulti* to remove the variables that contributed the least to model fit. We then ran final GLM models using only the variables selected by *glmulti*.

We evaluated the exploratory models using the true skill statistic (TSS; Allouche et al. 2006) and the testing AUC. AUC is a threshold-independent measure of accuracy while TSS is threshold-dependent (Komac et al. 2016): AUC is therefore more suited to evaluating the performance of continuous probability scores and TSS is more suited to evaluating binary predictions of presence/absence based on the threshold value (Allouche et al. 2006). TSS values range from -1 to 1, with 1 representing a perfectly accurate model and values less than 0

indicating a performance no better than random (Allouche et al. 2006). AUC estimates can be misleading when generating pseudo-absences from more distant areas, but we avoided this bias by restricting predictions to the known historic bridle shiner range (Lobo et al. 2008; Sutton et al. 2015).

2.2.5.5. Ensemble Models and Predictions. We chose an ensemble model approach in order to emphasize the trends emerging from the data while reducing the noise from individual model outputs (Araújo and New 2007). We selected the final set of predictor variables (n = 11) by comparing AUC and TSS values from exploratory analyses. We used this final set of variables and k-fold cross-validation (k = 10 folds or seeds) to train final RF, Maxent, and GLM models for the current and historic time periods. We tuned Maxent and RF hyperparameters a second time with the reduced set of predictor variables (*sensu* Vignali et al. 2020). We then calculated the probability that each cell of an output raster would be occupied by bridle shiners using the final, cross-validated models.

Raw probability scores generated by a model need to be rescaled by species prevalence in order to reflect habitat suitability (Jiménez-Valverde and Lobo 2007; Lobo et al. 2008). We accomplished this by calculating the mean threshold value (10 threshold values from 10-fold cross-validation) which maximized the sum of model sensitivity and specificity for each model (Jiménez-Valverde and Lobo 2007; Komac et al. 2016). All values below the threshold probability value are considered absences and all values above the threshold are considered presences. This threshold varies with each model and defines the value below which the combination of predictor variables is considered unsuitable habitat. Binarizing the rasters allowed us to quantify the predicted area of suitable habitat (Fourcade 2021). We then subtracted the historic presence/absence raster from the current presence/absence raster to determine the

overall change in predicted occupied area. To approximate only areas of shallow water, we created a 100-m lakeshore buffer and removed lake centers from the final ensemble model output. We then calculated the total area of suitable habitat in Maine and New Hampshire, along with the change in predicted occupied area for each state.

2.3. Results

2.3.1. Local Habitat Variables

We fit a 10-fold cross-validated GLM on 35 Maine habitat variables. We ranked variables according to their average variable importance, and eliminated eight lower-ranking variables that were highly correlated with higher-ranking variables. We ran all linear combinations of the selected 27 variables in *glmulti*, whose top model eliminated an additional 16 variables (Appendix G).

We then fit a final GLM with the 11 covariates retained by the top model. All of these variables influenced the probability of bridle shiner presence: the number of dams within 2-km of the site (Z = -2.04, p = 0.04), the proportion of deciduous forest within a site's HUC12 (Z = -2.04, p = 0.04), the proportion of floating vegetation (Z = -2.04, p = 0.04), the proportion of persistent emergent vegetation (Z = -2.03, p = 0.04), the total area of a site's HUC12 (Z = 2.03, p = 0.04), the Index of Ecological Integrity (Z = 2.03, p = 0.04), the proportion of complex-leaved submerged vegetation (Z = 2.03, p = 0.04), the proportion of mixed forest within a site's HUC12 (Z = 2.02, p = 0.04), the proportion of coniferous forest within a site's HUC12 (Z = -2.02, p = 0.04), the proportion of agricultural land in the site's upstream catchment (Z = -2.02, p = 0.04), and the proportion of deciduous forest in a site's upstream catchment (Z = -2.01, p = 0.04). The χ^2 analysis of deviance revealed that the proportion of persistent emergent vegetation ($p < 0.001$), the proportion of complex-leaved submerged vegetation ($p < 0.01$), the proportion of coniferous

forest in a HUC12 ($p = 0.01$), the total area of a site's HUC12 ($p = 0.01$), and the IEI ($p = 0.03$) all significantly improved model fit.

Therefore, bridle shiners in Maine were more likely to occur at sites with a higher proportion of complex-leaved submerged vegetation than floating or persistent emergent vegetation. We also found evidence that bridle shiners were more likely to occur in HUC12 sub-watersheds with more mixed forest than strictly deciduous or coniferous forest. We were more likely to find them in areas with a higher Index of Ecological Integrity and in areas with fewer dams. Sites with a higher proportion of agriculture and deciduous forest in their upstream drainage were less likely to support bridle shiners. Bridle shiners were also more likely to inhabit sites with larger HUC12 sub-watersheds.

2.3.2. Species Distribution Models

We modeled the historic and current range of the bridle shiner within Maine and New Hampshire. Of the six exploratory SDMs, the Maxent model of current bridle shiner presence had the highest AUC (0.92) and the GLM of current bridle shiner presence had the highest TSS (0.74; Table 2.3). We trained the six final individual models using the 11 variables retained by the current Maxent model: catchment position (*catchment*), soil clay content (*clay*), elevation (*elev*), proportion of Laurentian-Acadian Northern Pine(-Oak) Forest (*for.1924*), proportion of Appalachian (Hemlock-)Northern Hardwood Forest (*for.1927*), proportion of Central Appalachian Pine-Oak Rocky Woodland (*for.1930*), proportion of Boreal Jack Pine-Black Spruce Forest (*for.1980*), lithology (*lith*), soil pH (*pH*), soil silt content (*silt*), and slope (*slope*).

Table 2.3 AUC and TSS performance scores for the three exploratory species distribution modeling methods (generalized linear model [GLM], Maxent, and random forest [RF]) for the historic (1898-1999) and current (2000-2022) bridle shiner distribution. Exploratory models were generated using 24 predictor variables.

	trainAUC		testAUC		TSS	
	Historic	Current	Historic	Current	Historic	Current
GLM	0.886	0.897	0.843	0.911	0.573	0.742
Maxent	0.929	0.962	0.870	0.921	0.642	0.727
RF	1.000	1.000	0.907	0.881	0.690	0.628

While bridle shiner occurrences also overlap with Central Appalachian Dry Oak-Pine Forest (*for.1926*), only the historic (exploratory) Maxent model assigned this variable a permutation importance over zero (5.6% permutation importance). The exploratory *glmulti* of historic populations returned 29 models within 2 ΔAICc of the top model, with only two of these models including *for.1926* (Appendix H: Table H.1). The current *glmulti* returned 22 models within 2 ΔAICc of the top model, with only one of the models including *for.1926* (Appendix H: Table H.2). While not associated with bridle shiner occurrences in Maine and New Hampshire, this and other forest types may be predictive of bridle shiner presence in the central and southern portions of their range.

All model types performed better than random (AUC > 0.70, Baldwin 2009; TSS > 0.60, Komac et al. 2016) for both the historic and current time periods (Table 2.4), although the historic GLM had a TSS score slightly below 0.60 (TSS = 0.59), indicating only moderate support for this model (Landis and Koch 1977). All three model types performed better when evaluating current bridle shiner presence-absence data than when evaluating historic presence-only data (Table 2.4).

Table 2.4 AUC and TSS performance scores and threshold suitability values for the three species distribution modeling methods (generalized linear model [GLM], Maxent, and random forest [RF]) and resulting ensemble models of historic (1898-1999) and current (2000-2022) bridle shiner distribution. Final models were generated using 11 predictor variables.

	trainAUC		testAUC		TSS		Threshold	
	Historic	Current	Historic	Current	Historic	Current	Historic	Current
GLM	0.864	0.884	0.847	0.866	0.588	0.622	0.015	0.016
Maxent	0.929	0.964	0.875	0.921	0.633	0.745	0.127	0.141
RF	1.000	1.000	0.902	0.908	0.670	0.728	0.011	0.009
Ensemble	0.931	0.949	0.875	0.899	0.630	0.699	0.051	0.055

The ensemble model predicted a 51% reduction of suitable bridle shiner habitat between the historic (676-km^2) and current (331-km^2) time periods. Predicted habitat loss was more pronounced in Maine, where only 37.7% of historic bridle shiner habitat remains. The ensemble models also predicted a substantial loss of habitat (45.7%) in New Hampshire, but lost habitats were interspersed with predicted habitat gains (Figure 2.3). The majority of these "new" habitats were located in streams and rivers, but we did not see a corresponding trend in Maine. This could reflect differences in variable importance between the historic and current models: Appalachian (Hemlock-)Northern Hardwood Forest (*for.1927*), for example, was the primary factor influencing the historic random forest model (65.4% permutation importance) but not the current random forest model (12.3% permutation importance; Figure 2.4b). The variable importance of the Maxent models and GLMs did not vary as substantially over time.

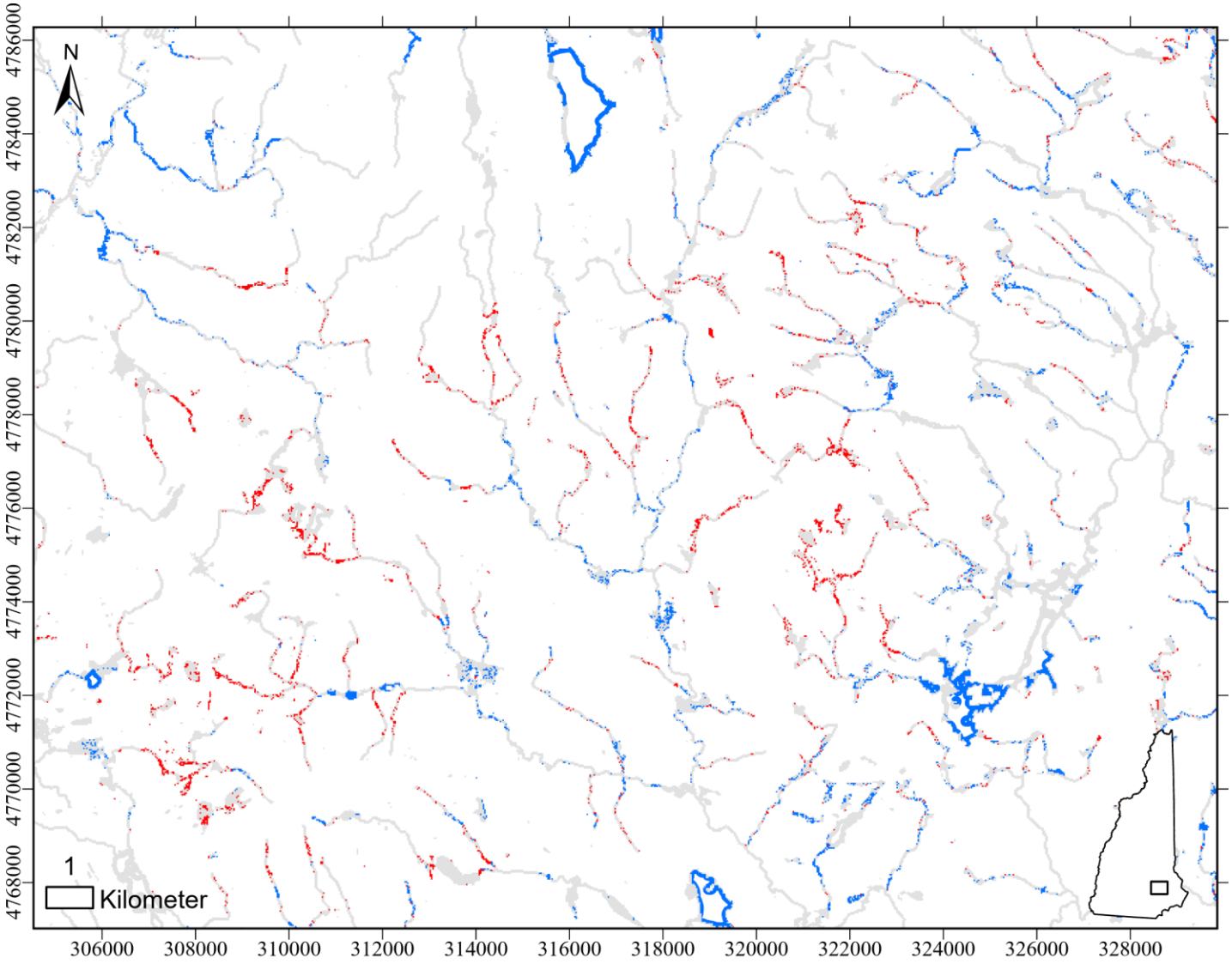


Figure 2.3 Ensemble model-predicted areas of bridle shiner habitat loss (blue), gain (red), and no change (gray) over southeastern New Hampshire.

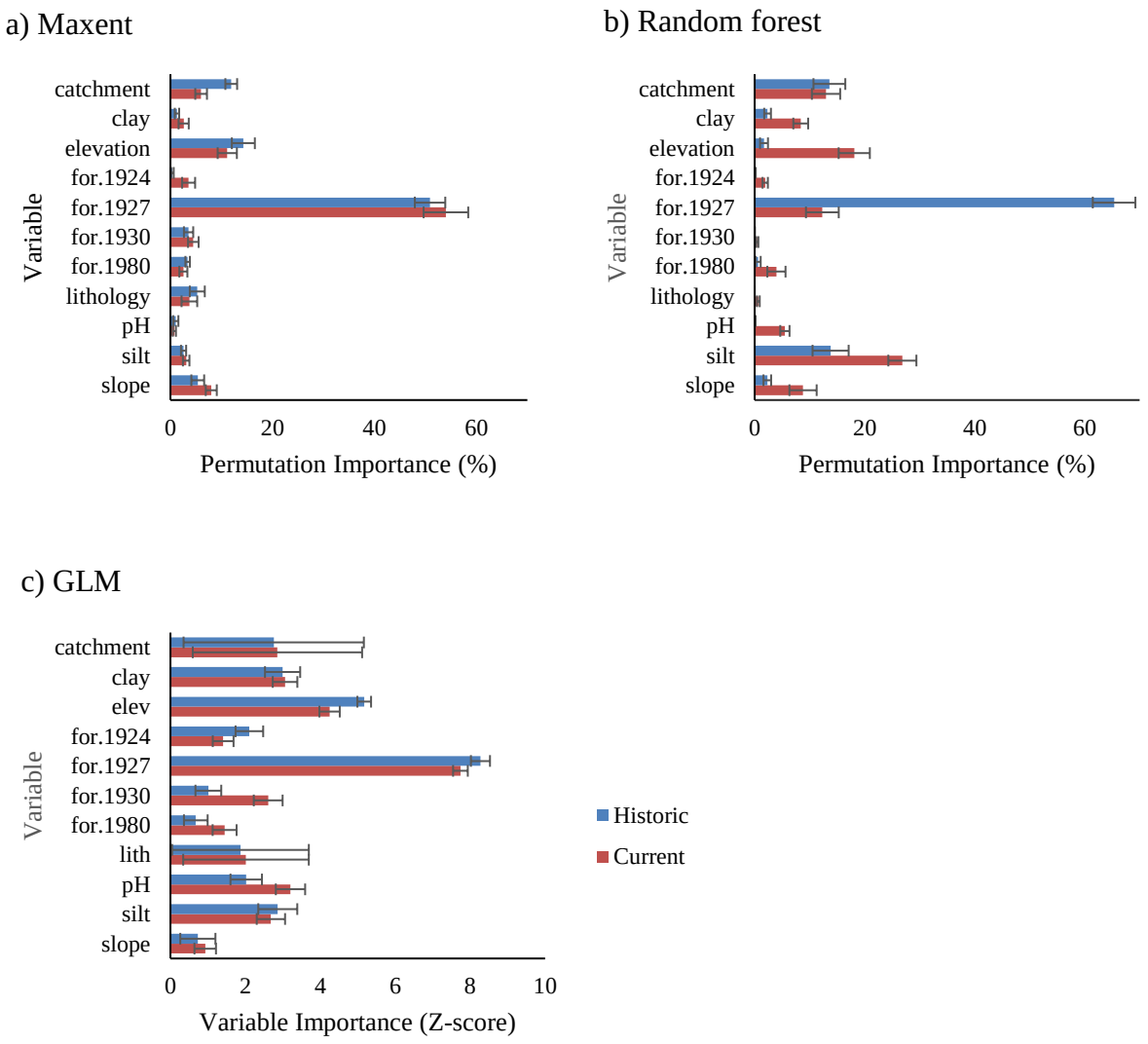


Figure 2.4 Permutation importance of the 11 environmental variables included in the final a) Maxent, b) random forest, and c) GLM models for both the historic (blue) and current (red) time periods.

Predicted suitable habitats in Maine and New Hampshire were once distributed throughout the Saco and Merrimack basins (HUC6) but are now relegated to the northeastern and western portions of the Saco basin and the eastern and central portions of the Merrimack (Figure 2.5a, b). Much of the once-suitable habitat in the central Saco and southeastern Merrimack is predicted to have been lost, and the limits of the bridle shiner range seem to be shifting westward and inland (Figure 2.5b, c). Bridle shiners were predicted to have historically occupied areas east

of the Saco basin (Lower Androscoggin HUC8), but only a small portion of this area remains in the current model (Figure 2.5c).

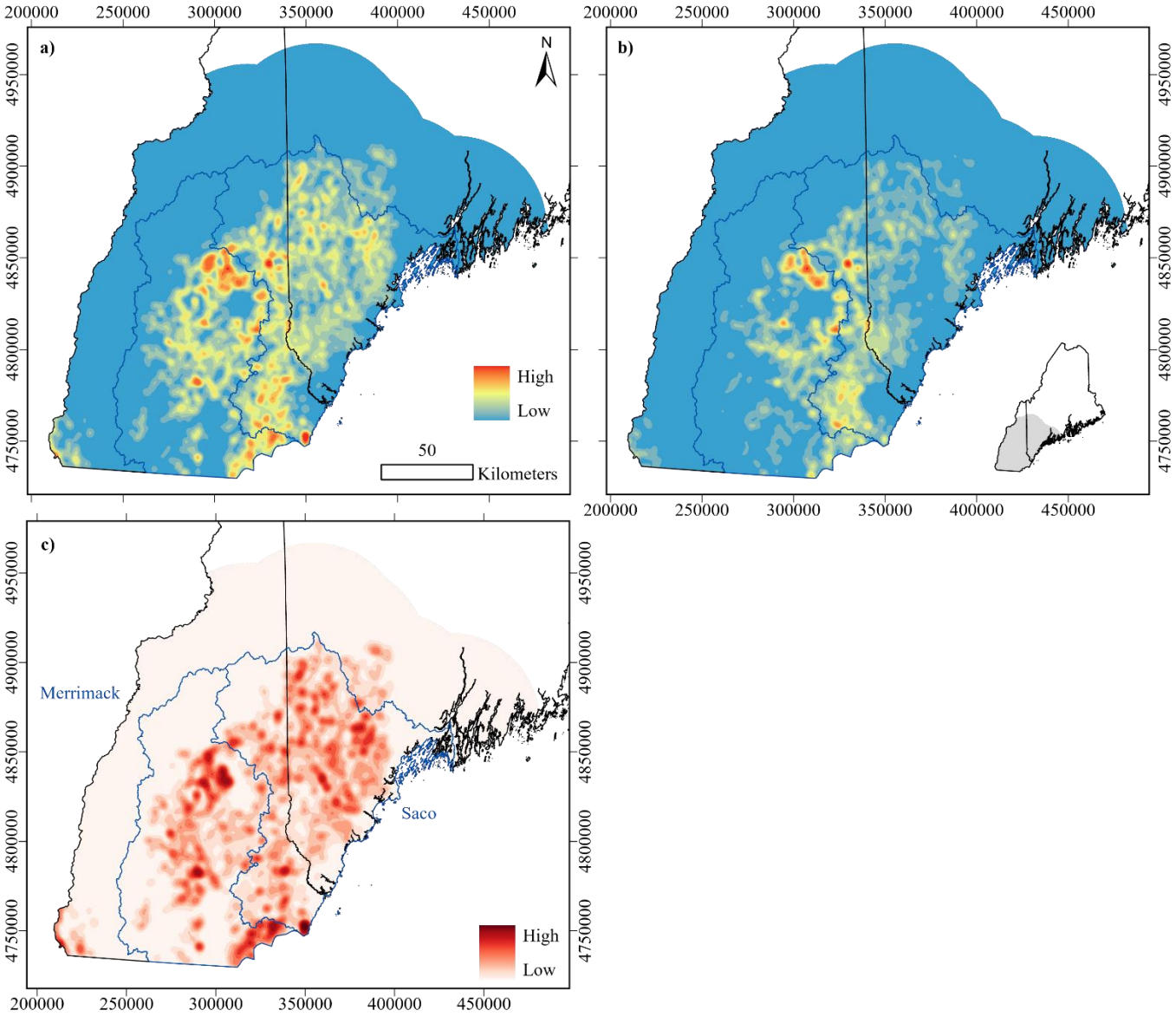


Figure 2.5 Kernel density of predicted suitable bridle shiner habitat in southeastern New Hampshire and southwestern Maine calculated using results of a) historic (1898-1999) and b) current (2000-2022) species distribution models and a 100-m radius. Predicted habitat loss over time (c) is also represented by a kernel density plot. Inset map shows modeled region in light gray, state boundaries are in black, and the Saco and Merrimack basin (HUC6) boundaries are shown in dark blue.

2.3.2.1. Variable Importance. We used the permutation importance of the final Maxent and RF models and the variable importance of the final GLMs to determine the effect size of the 11 covariates on bridle shiner presence (Figure 2.4). Additionally, we used the *Z*-values of the ten-fold, cross-validated GLMs to determine the directionality of the effects (Tables 2.5 and 2.6). The most influential variable in five of the six final models was Appalachian (Hemlock-)Northern Hardwood Forest (*for.1927*), with other forest types contributing less to the final model outputs (Figure 2.4). Bridle shiners were positively associated with this forest type and with Boreal Jack Pine-Black Spruce Forest (*for.1980*) and negatively associated with Laurentian-Acadian Northern Pine(-Oak) Forest (*for.1924*) and Central Appalachian Pine-Oak Rocky Woodland (*for.1930*; Tables 2.5 and 2.6).

Table 2.5 Cross-validated generalized linear model *Z*-values for each predictor variable and each level of categorical variable in the historic (1898-1999) bridle shiner species distribution model. Ten-fold cross-validation was achieved using ten random seeds.

	Seed 1	Seed 2	Seed 3	Seed 4	Seed 5	Seed 6	Seed 7	Seed 8	Seed 9	Seed 10	Average
(Intercept)	-14.06	-14.36	-14.33	-14.62	-14.52	-14.70	-14.75	-14.24	-14.27	-14.59	-14.44
catchment2	3.89	4.05	4.13	4.62	4.20	4.33	4.32	4.27	3.88	3.64	4.13
catchment3	4.34	4.30	3.93	6.07	4.65	5.24	5.00	4.84	4.85	4.95	4.82
catchment4	3.99	4.44	4.11	5.67	4.85	5.00	5.14	4.57	4.54	4.73	4.70
catchment5	-0.03	0.05	-0.02	-0.03	0.04	0.33	0.45	-0.03	0.09	0.08	0.09
catchment6	-0.01	-0.02	-0.03	-0.01	-0.03	-0.02	-0.02	-0.01	-0.03	-0.02	-0.02
clay	-2.86	-3.19	-3.73	-2.75	-3.09	-2.91	-3.22	-2.62	-2.73	-2.82	-2.99
elev	-5.50	-5.27	-5.46	-4.61	-5.31	-4.95	-4.96	-5.18	-5.36	-5.12	-5.17
for.1924	-2.10	-1.59	-2.07	-1.94	-1.79	-2.33	-2.16	-2.19	-2.49	-2.41	-2.11
for.1927	8.12	8.38	8.02	8.23	8.16	8.20	8.73	8.40	8.29	8.21	8.27
for.1930	-0.77	-1.53	-1.12	-1.06	-0.94	-0.65	-1.62	-0.99	0.32	-1.11	-0.95
for.1980	1.04	0.89	0.82	0.12	0.40	0.56	0.98	0.51	0.40	1.04	0.68
lith13	-0.46	0.05	-0.34	-1.33	-0.74	-1.30	-1.13	-0.57	-1.05	-0.27	-0.71
lith14	-2.75	-2.25	-2.16	-2.44	-2.72	-2.55	-2.14	-2.81	-2.82	-2.76	-2.54
lith19	-0.01	-0.02	-0.02	-0.01	-0.03	-0.02	-0.02	-0.01	-0.03	-0.02	-0.02
lith999	-3.87	-4.30	-3.48	-4.50	-4.42	-4.44	-4.20	-4.17	-4.33	-4.28	-4.20
pH	-2.09	-1.61	-2.03	-2.13	-1.60	-1.97	-2.91	-1.65	-1.94	-2.30	-2.02
silt	2.95	2.46	2.65	2.37	3.00	2.44	3.07	3.01	3.13	3.57	2.86
slope	-0.59	-0.76	-1.25	-0.40	-0.60	-1.03	-0.90	-0.68	-0.29	-0.81	-0.73

Table 2.5 Continued.

Note: NAHCS stream orders: catchment2 = 1b:Creek: >=3.861<38.61 sq.mi., catchment3 = 2:Small River: >= 38.61<200 sq.mi., catchment4 = 3a:Medium Tributary River: >=200<1000 sq.mi., catchment5 = 3b:Medium Mainstem River: >=1000<3861 sq.mi., catchment6 = 4:Large River: >=3861<9653 sq.mi., 5:Great River: >=9653 sq.mi.
LANDFIRE Biophysical Settings: for.1924 = Laurentian-Acadian Northern Pine(-Oak) Forest, for.1927 = Appalachian (Hemlock-)Northern Hardwood Forest, for.1930 = Central Appalachian Pine-Oak Rocky Woodland, for.1980 = Boreal Jack Pine-Black Spruce Forest
Lithology: lith999 = water, lith13 = glacial lake sediment fine, lith14 = glacial outwash coarse, lith19 = alluvium and coastal sediment fine

Table 2.6 Cross-validated generalized linear model *Z*-values for each predictor variable and each level of categorical variable in the current (2000-2022) bridle shiner species distribution model. Ten-fold cross-validation was achieved using ten random seeds.

	Seed 1	Seed 2	Seed 3	Seed 4	Seed 5	Seed 6	Seed 7	Seed 8	Seed 9	Seed 10	Average
(Intercept)	-14.20	-14.04	-13.38	-13.72	-13.95	-13.85	-13.91	-13.37	-13.83	-13.20	-13.74
catchment2	3.94	4.15	3.34	4.11	3.41	4.06	3.96	3.98	3.71	4.02	3.87
catchment3	5.48	5.14	4.91	5.34	4.70	5.08	4.98	5.67	4.91	5.37	5.16
catchment4	4.95	5.27	5.35	4.58	5.20	4.53	4.94	5.75	5.75	5.87	5.22
catchment5	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03
catchment6	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
clay	-3.09	-3.26	-2.77	-2.85	-3.25	-2.62	-4.18	-3.04	-2.49	-3.05	-3.06
elev	-3.90	-4.44	-4.40	-4.49	-4.23	-4.27	-4.28	-4.22	-4.01	-4.25	-4.25
for.1924	-2.01	-1.58	-1.00	-1.53	-1.51	-1.75	-1.22	-1.58	-1.05	-0.84	-1.41
for.1927	7.61	7.61	7.52	7.50	7.93	7.64	8.21	7.75	7.54	8.11	7.74
for.1930	-2.06	-2.71	-2.61	-2.43	-2.62	-2.59	-3.34	-2.23	-2.74	-2.76	-2.61
for.1980	0.82	1.38	1.70	1.34	1.03	1.71	1.77	1.51	1.72	1.48	1.45
lith13	-0.97	-0.22	-1.01	-0.79	-0.38	-0.58	-0.55	-0.76	-1.33	-0.88	-0.75
lith14	-2.36	-2.88	-3.03	-2.76	-2.86	-2.90	-2.22	-2.57	-3.11	-2.88	-2.76
lith19	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
lith999	-4.49	-4.27	-5.05	-4.24	-4.40	-4.29	-3.84	-4.52	-5.09	-5.32	-4.55
pH	-2.70	-3.29	-2.56	-3.25	-3.72	-2.64	-3.42	-3.65	-3.38	-3.43	-3.20
silt	1.88	2.96	2.39	2.25	3.50	2.30	3.29	2.86	2.35	3.02	2.68
slope	-1.05	-0.14	-0.95	-0.39	-1.68	-0.56	-1.34	-0.80	-1.29	-1.10	-0.93

Note: NAHCS stream orders: catchment2 = 1b:Creek: >=3.861<38.61 sq.mi., catchment3 = 2:Small River: >= 38.61<200 sq.mi., catchment4 = 3a:Medium Tributary River: >=200<1000 sq.mi., catchment5 = 3b:Medium Mainstem River: >=1000<3861 sq.mi., catchment6 = 4:Large River: >=3861<9653 sq.mi., 5:Great River: >=9653 sq.mi.
LANDFIRE Biophysical Settings: for.1924 = Laurentian-Acadian Northern Pine(-Oak) Forest, for.1927 = Appalachian (Hemlock-)Northern Hardwood Forest, for.1930 = Central Appalachian Pine-Oak Rocky Woodland, for.1980 = Boreal Jack Pine-Black Spruce Forest
Lithology: lith999 = water, lith13 = glacial lake sediment fine, lith14 = glacial outwash coarse, lith19 = alluvium and coastal sediment fine

Catchment position and terrain (elevation, slope) variables moderately influenced bridle shiner presence across model type and time period (Figure 2.4). Specifically, catchment position *Z*-values were strongly positive for creeks (*catchment2*), small rivers (*catchment3*), medium tributary rivers (*catchment4*), and lakes/ponds fed primarily by any of these categories (Tables

2.5 and 2.6). Bridle shiner presence was weakly positively associated with medium mainstem rivers (*catchment5*) during the historic period (Table 2.5), but weakly negatively associated with this variable in the current time period model (Table 2.6). Similarly, bridle shiner presence was weakly negatively associated with large rivers (*catchment6*) across both time periods. Bridle shiner presence was inversely related to elevation (*elev*) and weakly inversely related to slope (slope; Tables 2.5 and 2.6). This suggests that bridle shiners occupy flatter sites at lower elevations, and that the bridle shiner range in Maine and New Hampshire may be limited to the north and west by mountain ranges.

Soil and lithology characteristics were also moderately influential for the RF and GLM models (Figure 2.4b, c). Bridle shiners were less likely to inhabit areas of coarse glacial outwash and areas identified as water by the lithology (*lith*) dataset (Tables 2.5 and 2.6). This dataset only categorized the largest lakes as water, so this negative relationship is likely showing that bridle shiners are less likely to inhabit large lakes than streams, rivers, and ponds. This corroborates the catchment position findings because larger lakes tended to have a higher catchment position value than smaller streams and rivers. Soils with a higher clay content reduced the probability of bridle shiner occupancy, while soils with a higher silt content increased it (Tables 2.5 and 2.6). GLMs also showed an inverse relationship between bridle shiner presence and soil pH: bridle shiners were more likely to occupy areas that were more acidic than average (average pH = 4.7; Tables 2.5 and 2.6). This finding may suggest that bridle shiners are more acid-tolerant than other cyprinids, who are considered intolerant of waters below pH 5.2 (Laerm et al. 1980; Rahel and Magnuson 1983). Alternatively, these results may suggest that surrounding soil pH is not indicative of a waterbody's pH.

2.4. Discussion

Bridle shiners have historically occurred in eastern New Hampshire and southwestern Maine where freshwater systems are more heavily degraded and stressed by human population growth, urban and exurban development, and climate change. Several populations in New Hampshire have become extirpated due to herbicide use, shoreline habitat loss, lake drawdowns, and eutrophication (NHFGD 2015). Recent eDNA and seine surveys in Maine (Chapter 1) allowed us to model habitat selection within a waterbody. We found that bridle shiners are strongly associated with sites that have a higher proportion of complex-leaved, submerged aquatic vegetation and a lower proportion of floating and persistent emergent vegetation. Bridle shiners were also more likely to inhabit sites with less anthropogenic disturbance (higher Index of Ecological Integrity). Using the proportion of 2019 land cover surrounding each site, we determined that bridle shiners were more likely to inhabit areas surrounded by mixed forest and with less agricultural land in their upstream catchment.

We also found that bridle shiners in Maine were more likely to persist at sites with fewer nearby dams. As with other freshwater and diadromous species, bridle shiner populations have become increasingly fragmented by dams and undersized culvert construction (Cote et al. 2009), but the cumulative impact of these barriers to bridle shiner metapopulation dynamics is not known. Bridle shiners are known to utilize the headponds of artificial dams (Geneva et al. 2018; Pregler et al. 2019), and in several instances have become extirpated from such habitats after sudden water level drops or dam breaches (NHFGD 2015). Although artificial dam impoundments can provide habitat for cyprinids, these impoundments also support a higher relative abundance of large piscivores (Whittum et al. 2023). It is possible that these artificial

habitats are population sinks for bridle shiners because of the high risk of predation and sudden water level fluctuations.

Bridle shiner declines are likely due to the same factors that affect other minnow species, especially habitat loss and degradation. Bridle shiners are vulnerable to practices such as lake drawdowns and herbicide use because they live on the shoreline and require access to abundant vegetation (Pregler et al. 2019). Occupancy modeling has shown that bridle shiners can be reliably detected via seine net (Jensen and Vokoun 2013; Pregler et al. 2015) and environmental DNA (Chapter 1), so range-wide declines likely reflect true absences and extirpations rather than a failure to detect the species. These recent surveys in Maine and New Hampshire allowed us to model both their historic and current distribution in the region. We found that bridle shiner presence in these states is influenced by dominant forest type, catchment position, elevation, slope, soil composition, and lithology. Individual models had high model performance as determined by AUC and TSS statistics, and the overall ensemble models performed considerably better than random. Ensemble models predicted that only half (49%) of historic suitable habitat remains in this region, with losses in Maine being the most pronounced (62% decrease).

Our habitat suitability results mostly agreed with those of other bridle shiner habitat studies. Bridle shiners have been reported to prefer the still or slow-moving water of lakes, ponds, and low-gradient stream reaches (Jensen and Vokoun 2013; Pregler et al. 2019). While we were unable to include water velocity and stream gradient in our models, our final models considered point slope to be influential and inversely related to bridle shiner presence. This is additional evidence that bridle shiners inhabit areas with lower slopes such as lakes, ponds, and river backwaters.

Bridle shiners in Connecticut select reaches or habitat patches with unconsolidated bottoms, silty substrate, and abundant aquatic vegetation (Jensen and Vokoun 2013; Pregler et al. 2019). At the landscape scale, we found that areas with a higher soil silt content were significantly more likely to support bridle shiners, while areas with soil high in clay are less likely to support bridle shiners.

Pregler et al. (2019) also found that water body catchment position had a statistically significant effect on the probability of bridle shiner occurrence: water bodies in the headwaters of a catchment were more likely to support bridle shiners. Our SDMs provided further evidence that creeks (1b), small rivers (2), and medium tributary rivers (3a) were more likely to support bridle shiners than headwater streams (1a) and large rivers (positions 3b and 4). It is possible that bridle shiners in Maine and New Hampshire prefer water bodies further downstream in a catchment because of the cold temperatures in headwater streams. In Maine, headwater streams were colder than other stream orders and tended to have higher stream gradients.

Pregler et al. (2019) also found that bridle shiners in Connecticut were more likely to persist in areas of high forest cover and low impervious cover. We included the proportions of a suite of forest types in our models to add to their predictive power and found that Appalachian (Hemlock-)Northern Hardwood Forest cover was strongly associated with bridle shiner presence. The influence of this forest type may be specific to this region, as it does not extend much further south than New Hampshire. Other forest types, such as Central Appalachian Dry Oak-Pine Forest or Northern Atlantic Coastal Plain Hardwood Forest, may be more predictive of bridle shiner occurrence in central and southern portions of their range. We used LANDFIRE Biophysical Settings forest classifications because they are based on both the current biophysical environment and historical disturbance regimes (Rollins 2009; Blankenship et al. 2021), and so

were applicable to both of our modeling time scales (1898-1999 and 2000-2022). Measures of general forest composition, such as the 2019 NLCD mixed, coniferous, and deciduous forest classifications, may be more appropriate for modeling larger portions of the bridle shiner's current range.

Bridle shiners have been documented using the low salinity portions of estuaries in the southern part of their range (Cooper 1985). Our models suggest that the northeastern bridle shiner range is shifting away from the coast, so including a dataset of salinity gradients could be informative in future models. Similarly, a dataset of pH gradients across waterbodies could allow us to discern if this variable limits bridle shiner distribution. We attempted to model this using the pH of nearby soil as a proxy, which suggested that bridle shiners occupied more acidic areas. This is the opposite of what we would expect given the literature on other cyprinid species, including the closely-related blacknose shiner, which are generally intolerant of low pH (Laerm et al. 1980; Rahel and Magnuson 1983; Rahel 1984).

2.4.1. Study Limitations

One of the limitations of using background, or pseudo-absence, points in SDMs is the high degree of class overlap between presence and background variables: a portion of the randomly-generated points will occur in areas that have suitable habitat, and may even have undocumented populations of the species (Valavi et al. 2021). Generating several thousand background points is also necessary when characterizing the range of environmental conditions in the modeled area, which creates an imbalance between the number of presence and background points (Valavi et al. 2021). Using a large number of randomly-selected background points is recommended when using regression techniques and Maxent models (Barbet-Massin et al. 2012), but can result in overly complex ("overfit") RF models. While our final historic RF

models had the highest test AUCs of all three model types, their training AUCs were both 1.0, which is evidence of model overfitting.

Maxent modeling has also received scrutiny because it does not produce estimates of the probability of presence, but rather estimates an index of habitat suitability for each raster cell between 0 and 1 (Elith et al. 2011; Royle et al. 2012). Maxent also uses a complementary log-log ("cloglog") link function by default, which assumes that a species' presence or absence at nearby sites is independent (Phillips et al. 2017). This is not an appropriate assumption for bridle shiners, whose limited dispersal ability results in spatial autocorrelation of presence points (Phillips et al. 2017). Therefore, we used a logistic link function to predict habitat suitability on the probability scale so that Maxent outputs could be averaged with RF and GLM outputs.

Maxent typically uses presence-only data and randomly generates pseudo-absences. In general, presence-absence data is preferred to presence-only data because observed zeros are more informative than points with unknown occupancy (Royle et al. 2012). We used our observed absences in addition to randomly generated background points in all of our SDM models, including Maxent. All of the models run with presence-absence-background data performed better (higher AUC and TSS) than models run with only presence-background data, although we cannot say for certain that the inclusion of absence points was what improved model performance. Model performance was also likely improved because all of the lake and pond sites that were missing coordinates, and for which we used water body centroids, were included as absences in the current time period models. Including these points as presences may have reduced the performance of the historic models.

There are several environmental variables that are currently unavailable in spatial databases that could improve future iterations of our SDMs. First, lake and river bathymetry data

are not available for water bodies in Maine. Combining depth information with other currently unmapped features such as river and lake substrate, water velocity, and submerged vegetation density would provide detailed predictions of specific areas within water bodies where bridle shiners are likely to persist. This would eliminate the need to crop out portions of the final models to approximate shallow areas.

2.4.2. Conclusions

Locating additional bridle shiner populations in Maine and New Hampshire, especially at the periphery of their predicted range in Maine, will be critical to preventing further declines. These peripheral populations merit high conservation priority because of the unique evolutionary pressures they have faced compared with conspecifics at the core of their range (Taylor et al. 2003; Haak et al. 2010), and because they are at the leading edge of the species' potential northward expansion in response to climate change (Gibson et al. 2009). Using spatially explicit habitat models to target survey areas can result in greater survey efficiency over large areas (Guisan et al. 2006). Our local and regional models can be used to focus surveys on areas across Maine and New Hampshire with high predicted habitat suitability. In addition to guiding the search for undiscovered bridle shiner populations, managers may also use these models to search for suitable reintroduction or assisted migration sites or to inform habitat restoration efforts.

BIBLIOGRAPHY

Adams, C. C., and T. L. Hankinson. 1928. The ecology and economics of Oneida Lake fish. Roosevelt Wild Life Annals 1(3–4):235–548.

Agersnap, S., E. E. Sigsgaard, M. R. Jensen, M. D. P. Avila, H. Carl, P. R. Møller, S. L. Krøs, S. W. Knudsen, M. S. Wisz, and P. F. Thomsen. 2022. A national scale "BioBlitz" using citizen science and eDNA metabarcoding for monitoring coastal marine fish. Frontiers in Marine Science 9:1–17.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6):716–723.

Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43(6):1223–1232.

Angermeier, P. L. 1995. Ecological attributes of extinction-prone species: loss of freshwater fishes of Virginia. Conservation Biology 9(1):143–158.

Araújo, M. B., and M. New. 2007. Ensemble forecasting of species distributions. Trends in Ecology & Evolution 22(1):42–47.

Baldwin, R. 2009. Use of maximum entropy modeling in wildlife research. Entropy 11:854–866.

Barbet-Massin, M., F. Jiguet, C. Albert, and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution 3:327–338.

Barnes, M. A., C. R. Turner, C. L. Jerde, M. A. Renshaw, W. L. Chadderton, and D. M. Lodge. 2014. Environmental conditions influence eDNA persistence in aquatic systems. Environmental Science & Technology 48:1819–1827.

Biggs, J., N. Ewald, A. Valentini, C. Gaboriaud, T. Dejean, R. A. Griffiths, J. Foster, J. W. Wilkinson, A. Arnell, P. Brotherton, P. Williams, and F. Dunn. 2015. Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). Biological Conservation 183:19–28.

Blankenship, K., R. Swaty, K. R. Hall, S. Hagen, K. Pohl, A. Shlisky Hunt, J. Patton, L. Frid, and J. Smith. 2021. Vegetation dynamics models: a comprehensive set for natural resource assessment and planning in the United States. Ecosphere 12(4):e03484.

Bonar, S. A., M. Divens, and B. Bolding. 1997. Methods for sampling the distribution and abundance of bull trout/Dolly Varden. Washington Dept. of Fish and Wildlife, Fish Management Program, Inland Fisheries Investigations, Resources Assessment Division, Research report RAD97-05, Olympia, WA.

Breiman, L. 2001. Random forests. Machine Learning 45(1):5–32.

Bunn, S. E., and A. H. Arthington. 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. Environmental Management 30(4):492–507.

Calcagno, V. 2020. glmulti: model selection and multimodel inference made easy. R package version 1.0.8. https://CRAN.R-project.org/package=glmulti.

Carlos-Júnior, L. A., J. C. Creed, R. Marrs, R. J. Lewis, T. P. Moulton, R. Feijó-Lima, and M. Spencer. 2020. Generalized linear models outperform commonly used canonical analysis in estimating spatial structure of presence/absence data. PeerJ 8:e9777.

Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18(1):117–143.

Collins, R. A., J. Bakker, O. S. Wangensteen, A. Z. Soto, L. Corrigan, D. W. Sims, M. J. Genner, and S. Mariani. 2019. Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods in Ecology and Evolution 10(11):1985–2001.

Committee on the Status of Endangered Wildlife in Canada [COSEWIC]. 2013. COSEWIC assessment and status report on the bridle shiner (*Notropis bifrenatus*) in Canada. CW69-14/671-2013E-PDF, Ontario, Canada.

Cooper, E. L., editor. 1985. Chapter 3 - Fishes. Pages 169–256 Species of Special Concern in Pennsylvania. Carnagie Museum of Natural History, Pittsburgh, PA.

Cooper, G. P. 1939. A biological survey of thirty-one lakes and ponds of the Upper Saco River and Sebago Lake drainage systems in Maine. Maine Department of Inland Fisheries and Game, Fish Survey Report 2.

Cote, D., D. G. Kehler, C. Bourne, and Y. F. Wiersma. 2009. A new measure of longitudinal connectivity for stream networks. Landscape Ecology 24(1):101–113.

Coulston, J. W., G. G. Moisen, B. T. Wilson, M. V. Finco, W. B. Cohen, and C. K. Brewer. 2012. Modeling percent tree canopy cover: a pilot study. Photogrammetric Engineering & Remote Sensing 78(7):715–727.

Cowardin, L. M., V. Carter, F. C. Golet, and E. T. LaRoe. 1979. Classification of wetlands and deepwater habitats of the United States. Page Classification of Wetlands and Deepwater Habitats of the United States. U.S. Department of the Interior, Fish and Wildlife Service Washington, D.C. USA.

Deeds, J., A. Amirbahman, S. A. Norton, and L. C. Bacon. 2020. A hydrogeomorphic and condition classification for Maine, USA, lakes. Lake and Reservoir Management 36(2):122–138.

Deiner, K., and F. Altermatt. 2014. Transport distance of invertebrate environmental DNA in a natural river. PLoS ONE 9(2):e88786.

Deiner, K., E. A. Fronhofer, E. Mächler, J.-C. Walser, and F. Altermatt. 2016. Environmental DNA reveals that rivers are conveyer belts of biodiversity information. Nature Communications 7(1):12544.

Dejean, T., A. Valentini, A. Duparc, S. Pellier-Cuit, F. Pompanon, P. Taberlet, and C. Miaud. 2011. Persistence of environmental DNA in freshwater ecosystems. PLoS ONE 6(8):e23398.

Dewitz, J., and USGS. 2021, June 4. National Land Cover Database (NLCD) 2019 Products. Raster, ScienceBase.

Doering, P. H., C. T. Roman, L. L. Beatty, A. A. Keller, and C. A. Oviatt. 1995. Water quality and habitat evaluation of Bass Harbor Marsh Acadia National Park, Maine. Page 213. National Park Service, New England System Support Office (NESO), NPS/NESORNR/NRTR/95-31, Boston, MA.

Dorazio, R. M., and R. A. Erickson. 2018. ednaoccupancy: An R package for multiscale occupancy modelling of environmental DNA data. Molecular Ecology Resources 18(2):368–380.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão, T. Münkemüller, C. McClean, P. E. Osborne, B. Reineking, B. Schröder, A. K. Skidmore, D. Zurell, and S. Lautenbach. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36(1):27–46.

Dudgeon, D., A. H. Arthington, M. O. Gessner, Z.-I. Kawabata, D. J. Knowler, C. Lévêque, R. J. Naiman, A.-H. Prieur-Richard, D. Soto, M. L. J. Stiassny, and C. A. Sullivan. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews 81(2):163–182.

Dumelle, M., T. Kincaid, A. R. Olsen, and M. Weber. 2023. spsurvey: spatial sampling design and analysis in R. Journal of Statistical Software 105:1–29.

Eichmiller, J. J., P. G. Bajer, and P. W. Sorensen. 2014. The relationship between the distribution of common carp and their environmental DNA in a small lake. PLoS ONE 9(11):1–8.

Eichmiller, J. J., L. M. Miller, and P. W. Sorensen. 2016. Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. Molecular Ecology Resources 16(1):56–68.

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17(1):43–57.

Everhart, W. H. 2002. Fishes of Maine. Maine Dept. of Inland Fisheries and Wildlife, Augusta, Maine.

Fagan, W. F. 2002. Connectivity, fragmentation, and extinction risk in dendritic metapopulations. Ecology 83(12):3243–3249.

Ficetola, G. F., J. Pansu, A. Bonin, E. Coissac, C. Giguet-Covex, M. De Barba, L. Gielly, C. M. Lopes, F. Boyer, F. Pompanon, G. Rayé, and P. Taberlet. 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular Ecology Resources 15(3):543–556.

Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24(1):38–49.

Fiske, I. J., and R. B. Chandler. 2011. unmarked: an R Package for fitting hierarchical models of wildlife occurrence and abundance. Journal of Statistical Software 43(10).

Fourcade, Y. 2021. Fine-tuning niche models matters in invasion ecology. A lesson from the land planarian *Obama nungara*. Ecological Modelling 457:109686.

Gallagher, M. 2010a, January 8. Stream fish surveys (historical). http://www.gulfofmaine.org/kb/2.0/record.html?recordid=9233&save_record=1&search=quick&f_records_per_page=25&f_searchphrase=&f_documents_or_data=Data&f_records_per_page=25&submit=Search&sort=Author&f_start_this_page=0.

Gallagher, M. 2010b, January 8. Stream fish surveys (recent). http://www.gulfofmaine.org/kb/2.0/record.html?recordid=9234.

Geneva, A. J., A. M. Kreit, S. Neiffer, S. Tsang, and R. J. Horwitz. 2018. Regional population structure of the endangered Bridle Shiner (*Notropis bifrenatus*). Conservation Genetics 19(5):1039–1053.

Gibson, S. Y., R. C. Van Der Marel, and B. M. Starzomski. 2009. Climate change and conservation of leading-edge peripheral populations. Conservation Biology 23(6):1369–1373.

Goforth, R. R., and J. W. Foltz. 1998. Movements of the yellowfin shiner, *Notropis lutipinnis*. Ecology of Freshwater Fish 7(2):49–55.

Goldberg, C. S., K. M. Strickler, and A. K. Fremier. 2018. Degradation and dispersion limit environmental DNA detection of rare amphibians in wetlands: Increasing efficacy of sampling designs. Science of The Total Environment 633:695–703.

Gray, S., L. McDonnell, N. Mandrak, and L. Chapman. 2016. Species-specific effects of turbidity on the physiology of imperiled blackline shiners *Notropis* spp. in the Laurentian Great Lakes. Endangered Species Research 31:271–277.

Green, R. H., and R. C. Young. 1993. Sampling to detect rare species. Ecological Applications 3(2):351–356.

Greenwell, B., M., and B. Boehmke C. 2020. Variable importance plots — an introduction to the vip package. The R Journal 12(1):343.

Grieger, R. 2019, October 31. RPubs - NMDS ordination plotting. https://www.rpubs.com/RGrieger/545184.

Guisan, A., O. Broennimann, R. Engler, M. Vust, N. G. Yoccoz, A. Lehmann, and N. E. Zimmermann. 2006. Using niche-based models to improve the sampling of rare species. Conservation Biology 20(2):501–511.

Haak, A. L., J. E. Williams, H. M. Neville, D. C. Dauwalter, and W. T. Colyer. 2010. Conserving peripheral trout populations: the values and risks of life on the edge. Fisheries 35(11):530–549.

Hammerson, G. 2021, March 5. *Notropis bifrenatus*. NatureServe Explorer. https://explorer.natureserve.org/Taxon/ELEMENT_GLOBAL.2.100562/Notropis_bifrenatus.

Han, S., H. Kim, and Y.-S. Lee. 2020. Double random forest. Machine Learning 109(8):1569–1586.

Harrington, R. W. 1948a. The food of the bridled shiner, *Notropis bifrenatus* (Cope). The American Midland Naturalist 40(2):353–361.

Harrington, R. W. 1948b. The life cycle and fertility of the Bridled Shiner, *Notropis bifrenatus* (Cope). The American Midland Naturalist 39(1):83–92.

Hengl, T., M. Nussbaum, M. N. Wright, G. B. M. Heuvelink, and B. Gräler. 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6:e5518.

Hijmans, R. J. 2023. terra: spatial data analysis. R package version 1.7-29. https://CRAN.R-project.org/package=terra.

Hijmans, R. J., S. J. Phillips, J. Leathwick, and J. Elith. 2022. dismo: species distribution modeling. R package version 1.3-9. https://CRAN.R-project.org/package=dismo.

Hinlo, R., D. Gleeson, M. Lintermans, and E. Furlan. 2017. Methods to maximise recovery of environmental DNA from water samples. PLOS ONE 12(6):e0179251.

Hinlo, R., M. Lintermans, D. Gleeson, B. Broadhurst, and E. Furlan. 2018. Performance of eDNA assays to detect and quantify an elusive benthic fish in upland streams. Biological Invasions 20(11):3079–3093.

Jelks, H. L., S. J. Walsh, N. M. Burkhead, S. Contreras-Balderas, E. Diaz-Pardo, D. A. Hendrickson, J. Lyons, N. E. Mandrak, F. McCormick, J. S. Nelson, S. P. Platania, B. A. Porter, C. B. Renaud, J. J. Schmitter-Soto, E. B. Taylor, and M. L. Warren. 2008.

Conservation status of imperiled North American freshwater and diadromous fishes. Fisheries 33(8):372–407.

Jensen, T., and J. C. Vokoun. 2013. Using multistate occupancy estimation to model habitat use in difficult-to-sample watersheds: bridle shiner in a low-gradient swampy stream. Canadian Journal of Fisheries and Aquatic Sciences 70(10):1429–1437.

Jerde, C. L., A. R. Mahon, W. L. Chadderton, and D. M. Lodge. 2011. "Sight-unseen" detection of rare aquatic species using environmental DNA: eDNA surveillance of rare aquatic species. Conservation Letters 4(2):150–157.

Jiménez-Valverde, A., and J. M. Lobo. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. Acta Oecologica 31(3):361–369.

Johnston, C. E. 2000. Movement patterns of imperiled blue shiners (Pisces: Cyprinidae) among habitat patches. Ecology of Freshwater Fish 9(3):170–176.

Kaky, E., V. Nolan, A. Alatawi, and F. Gilbert. 2020. A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: A case study with Egyptian medicinal plants. Ecological Informatics 60:101150.

Kendall, W. C. 1914. An annotated catalogue of the fishes of Maine. Pages 1–216. Portland Society of Natural History, Portland, Me.

Kéry, M., and J. A. Royle. 2016. Applied hierarchical modeling in ecology: analysis of distribution, abundance and species richness in R and BUGS. Academic Press/Elsevier, Amsterdam; Boston.

Kilian, J. V., R. L. Raesly, S. A. Stranko, A. J. Becker, and E. Durell. 2011. Extirpation of the Bridle Shiner (*Notropis bifrenatus*) from Maryland. Northeastern Naturalist 18(2):236–242.

Komac, B., P. Esteban, L. Trapero, and R. Caritg. 2016. Modelization of the current and future habitat suitability of *Rhododendron ferrugineum* using potential snow accumulation. PloS one 11:e0147324.

Lacoursière-Roussel, A., G. Côté, V. Leclerc, and L. Bernatchez. 2016a. Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. Journal of Applied Ecology 53(4):1148–1157.

Lacoursière-Roussel, A., M. Rosabal, and L. Bernatchez. 2016b. Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. Molecular Ecology Resources 16(6):1401–1414.

Laerm, J., B. J. Freeman, L. J. Vitt, J. M. Meyers, and L. Logan. 1980. Vertebrates of the Okefenokee Swamp. Brimleyana 4:47–73.

Lahoz-Monfort, J. J., G. Guillera-Arroita, and R. Tingley. 2016. Statistical approaches to account for false-positive errors in environmental DNA samples. Molecular Ecology Resources 16(3):673–685.

Lamothe, K. A., and D. A. R. Drake. 2020. Habitat associations of the Threatened pugnose minnow (*Opsopoeodus emiliae*) at the northern edge of the species range. Ecology of Freshwater Fish 29(2):289–298.

Landis, J. R., and G. G. Koch. 1977. The measurement of observer agreement for categorical data. Biometrics 33(1):159–174.

Larson, E. R., B. M. Graham, R. Achury, J. J. Coon, M. K. Daniels, D. K. Gambrell, K. L. Jonasen, G. D. King, N. LaRacuente, T. I. Perrin-Stowe, E. M. Reed, C. J. Rice, S. A. Ruzi, M. W. Thairu, J. C. Wilson, and A. V. Suarez. 2020. From eDNA to citizen science: emerging tools for the early detection of invasive species. Frontiers in Ecology and the Environment 18(4):194–202.

Lesica, P., and F. W. Allendorf. 1995. When are peripheral populations valuable for conservation? Conservation Biology 9(4):753–760.

Lobo, J. M., A. Jiménez-Valverde, and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17(2):145–151.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. Andrew Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83(8):2248–2255.

Maine Dept. of Inland Fisheries and Wildlife. 2015. Maine's Wildlife Action Plan. Maine Department of Inland Fisheries and Wildlife, SWG Report, Augusta, ME.

Maine Dept. of Inland Fisheries and Wildlife. 2021. Species of Special Concern. https://www.maine.gov/ifw/fish-wildlife/wildlife/endangered-threatened-species/special-concern.html.

Marmion, M., M. Parviainen, M. Luoto, R. K. Heikkinen, and W. Thuiller. 2009. Evaluation of consensus methods in predictive species distribution modelling. Diversity and Distributions 15(1):59–69.

McArdle, B. H. 1990. When are rare species not there? Oikos 57(2):276.

McColl-Gausden, E. F., A. R. Weeks, R. A. Coleman, K. L. Robinson, S. Song, T. A. Raadik, and R. Tingley. 2020. Multispecies models reveal that eDNA metabarcoding is more sensitive than backpack electrofishing for conducting fish surveys in freshwater streams. Molecular Ecology 00:1–16.

McGarigal, K., B. W. Compton, E. B. Plunkett, W. V. DeLuca, J. Grand, E. Ene, and S. D. Jackson. 2018. A landscape index of ecological integrity to inform landscape conservation. Landscape Ecology 33(7):1029–1048.

Mize, E. L., R. A. Erickson, C. M. Merkes, N. Berndt, K. Bockrath, J. Credico, N. Grueneis, J. Merry, K. Mosel, M. Tuttle-Lau, K. V. Ruden, Z. Woiak, J. J. Amberg, K. Baerwaldt, S. Finney, and E. Monroe. 2019. Refinement of eDNA as an early monitoring tool at the landscape-level: study design considerations. Ecological Applications 29(6):e01951.

Mordecai, R. S., B. J. Mattsson, C. J. Tzilkowski, and R. J. Cooper. 2011. Addressing challenges when studying mobile or episodic species: hierarchical Bayes estimation of occupancy and use. Journal of Applied Ecology 48(1):56–66.

Nester, G. M., M. J. Heydenrych, T. E. Berry, Z. Richards, J. Wasserman, N. E. White, M. De Brauwer, M. Bunce, M. Takahashi, and L. Claassens. 2023. Characterizing the distribution of the critically endangered estuarine pipefish (*Syngnathus watermeyeri*) across its range using environmental DNA. Environmental DNA 5(1):132–145.

New Hampshire Fish and Game Department. 2015. New Hampshire Wildlife Action Plan. New Hampshire Fish and Game Department, SWG Report, Concord, NH.

Nichols, J. D., L. L. Bailey, A. F. O'Connell Jr., N. W. Talancy, E. H. Campbell Grant, A. T. Gilbert, E. M. Annand, T. P. Husband, and J. E. Hines. 2008. Multi-scale occupancy estimation and modelling using multiple detection methods. Journal of Applied Ecology 45(5):1321–1329.

Nohner, J. K., and J. S. Diana. 2015. Muskellunge spawning site selection in northern Wisconsin lakes and a GIS-based predictive habitat model. North American Journal of Fisheries Management 35(1):141–157.

O'Brien, J. 2023. gdalUtilities: Wrappers for "GDAL" Utilities Executables. R.

Oksanen, J., G. L. Simpson, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, P. Solymos, M. H. H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, G. Carvalho, M. Chirico, M. De Caceres, S. Durand, H. B. A. Evangelista, R. FitzJohn, M. Friendly, B. Furneaux, G. Hannigan, M. O. Hill, L. Lahti, D. McGlinn, M.-H. Ouellette, E. Ribeiro Cunha, T. Smith, A. Stier, C. J. F. ter Braak, and J. Weedon. 2022. vegan: Community ecology package. R package version 2.6-4. https://CRAN.R-project.org/package=vegan.

Olden, J. D., Z. S. Hogan, and M. J. V. Zanden. 2007. Small fish, big fish, red fish, blue fish: size-biased extinction risk of the world's freshwater and marine fishes. Global Ecology and Biogeography 16(6):694–701.

Olivero, A. P., and M. G. Anderson. 2008. Northeast Aquatic Habitat Classification System. The Nature Conservancy, Boston, MA.

Page, L. M., and B. M. Burr. 2011. Peterson field guide to freshwater fishes of North America north of Mexico, 2nd edition. Houghton Mifflin Harcourt Publishing Company, New York, NY.

Paul, M. J., and J. L. Meyer. 2001. Streams in the urban landscape. Annual Review of Ecology and Systematics 32:333–365.

Phillips, S. J., R. P. Anderson, M. Dudík, R. E. Schapire, and M. E. Blair. 2017. Opening the black box: an open-source release of Maxent. Ecography 40(7):887–893.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190(3):231–259.

Plough, L. V., M. B. Ogburn, C. L. Fitzgerald, R. Geranio, G. A. Marafino, and K. D. Richie. 2018. Environmental DNA analysis of river herring in Chesapeake Bay: A powerful tool for monitoring threatened keystone species. PLOS ONE 13(11):e0205578.

Poggio, L., L. M. de Sousa, N. H. Batjes, G. B. M. Heuvelink, B. Kempen, E. Ribeiro, and D. Rossiter. 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL 7(1):217–240.

Pregler, K. C., N. Hagstrom, E. T. Schultz, and J. C. Vokoun. 2019. Landscape factors predict local extirpation in an imperilled minnow species, the bridle shiner (*Notropis bifrenatus*). Aquatic Conservation: Marine and Freshwater Ecosystems 29(8):1227–1237.

Pregler, K. C., J. C. Vokoun, T. Jensen, and N. Hagstrom. 2015. Using multimethod occupancy estimation models to quantify gear differences in detection probabilities: is backpack electrofishing missing occurrences for a species of concern? Transactions of the American Fisheries Society 144(1):89–95.

Quinn, T. P., I. Erb, M. F. Richardson, and T. M. Crowley. 2018. Understanding sequencing data as compositions: an outlook and review. Bioinformatics 34(16):2870–2878.

R Core Team. 2022. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

R Core Team. 2023. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rahel, F. J. 1984. Factors structuring fish assemblages along a bog lake successional gradient. Ecology 65(4):1276–1289.

Rahel, F. J., and J. J. Magnuson. 1983. Low pH and the Absence of Fish Species in Naturally Acidic Wisconsin Lakes: Inferences for Cultural Acidification. Canadian Journal of Fisheries and Aquatic Sciences 40(1):3–9.

Riaz, M., M. Kuemmerlen, C. Wittwer, B. Cocchiararo, I. Khaliq, M. Pfenninger, and C. Nowak. 2020. Combining environmental DNA and species distribution modeling to evaluate reintroduction success of a freshwater fish. Ecological Applications 30(2):e02034.

Robinson, A. T., Y. M. Paroz, M. J. Clement, T. W. Franklin, J. C. Dysthe, M. K. Young, K. S. McKelvey, and K. J. Carim. 2019. Environmental DNA sampling of small-bodied minnows: performance relative to location, species, and traditional sampling. North American Journal of Fisheries Management 39(5):1073–1085.

Rollins, M. G. 2009. LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment. International Journal of Wildland Fire 18(3):235–249.

Rourke, M. L., A. M. Fowler, J. M. Hughes, M. K. Broadhurst, J. D. DiBattista, S. Fielder, J. W. Walburn, and E. M. Furlan. 2021. Environmental DNA (eDNA) as a tool for assessing fish biomass: A review of approaches and future considerations for resource surveys. Environmental DNA 00:1–25.

Roussel, J.-M., J.-M. Paillisson, A. Tréguier, and E. Petit. 2015. The downside of eDNA as a survey tool in water bodies. Journal of Applied Ecology 52(4):823–826.

Royle, J. A., R. B. Chandler, C. Yackulic, and J. D. Nichols. 2012. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. Methods in Ecology and Evolution 3(3):545–554.

Scheuerell, M. D., and D. E. Schindler. 2004. Changes in the spatial distribution of fishes in lakes along a residential development gradient. Ecosystems 7(1):98–106.

Schmidt, B. R., M. Kéry, S. Ursenbacher, O. J. Hyman, and J. P. Collins. 2013. Site occupancy models in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. Methods in Ecology and Evolution 4(7):646–653.

Soranno, P. A., K. S. Cheruvelil, T. Wagner, K. E. Webster, and M. T. Bremigan. 2015. Effects of land use on lake nutrients: the importance of scale, hydrologic connectivity, and region. PLOS ONE 10(8):e0135454.

de Souza, L. S., J. C. Godwin, M. A. Renshaw, and E. Larson. 2016. Environmental DNA (eDNA) detection probability is influenced by seasonal activity of organisms. PLOS ONE 11(10):e0165273.

Stone, J., B. C. Lê, and J. R. Moring. 2001. Freshwater fishes of Acadia National Park, Mount Desert Island, Maine. Northeastern Naturalist 8(3):311–318.

Strahler, A. N. 1957. Quantitative analysis of watershed geomorphology. Transactions of the American Geophysical Union 38(6):913–920.

Sutherland, J. W., S. A. Norton, J. W. Short, and C. Navitsky. 2018. Modeling salinization and recovery of road salt-impacted lakes in temperate regions based on long-term monitoring

of Lake George, New York (USA) and its drainage basin. Science of The Total Environment 637–638:282–294.

Sutton, W. B., K. Barrett, A. T. Moody, C. S. Loftin, P. G. DeMaynadier, and P. Nanjappa. 2015. Predicted changes in climatic niche and climate refugia of conservation priority salamander species in the northeastern United States. Forests 6(1):1–26.

Takahara, T., T. Minamoto, H. Yamanaka, H. Doi, and Z. Kawabata. 2012. Estimation of fish biomass using environmental DNA. PLoS ONE 7(4):e35868.

Takahashi, M., M. Saccò, J. H. Kestel, G. Nester, M. A. Campbell, M. van der Heyde, M. J. Heydenrych, D. J. Juszkiewicz, P. Nevill, K. L. Dawkins, C. Bessey, K. Fernandes, H. Miller, M. Power, M. Mousavi-Derazmahalleh, J. P. Newton, N. E. White, Z. T. Richards, and M. E. Allentoft. 2023. Aquatic environmental DNA: A review of the macro-organismal biomonitoring revolution. Science of The Total Environment 873:162322.

Taylor, E. B., M. D. Stamford, and J. S. Baxter. 2003. Population subdivision in westslope cutthroat trout (*Oncorhynchus clarki lewisi*) at the northern periphery of its range: evolutionary inferences and conservation implications. Molecular Ecology 12(10):2609–2622.

Thalinger, B., D. Kirschner, Y. Pütz, C. Moritz, R. Schwarzenberger, J. Wanzenböck, and M. Traugott. 2021. Lateral and longitudinal fish environmental DNA distribution in dynamic riverine habitats. Environmental DNA 3(1):305–318.

Theobald, D. M., D. Harrison-Atlas, W. B. Monahan, and C. M. Albano. 2015. Ecologically-relevant maps of landforms and physiographic diversity for climate adaptation planning. PLOS ONE 10(12):e0143619.

Thomsen, P. F., J. Kielgast, L. L. Iversen, P. R. Møller, M. Rasmussen, and E. Willerslev. 2012. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLOS ONE 7(8):e41732.

Turner, C. R., M. A. Barnes, C. C. Y. Xu, S. E. Jones, C. L. Jerde, and D. M. Lodge. 2014. Particle size distribution and optimal capture of aqueous macrobial eDNA. Methods in Ecology and Evolution 5(7):676–684.

U.S. Environmental Protection Agency [USEPA]. 2016. Region 1 Maine Lakes Data Sets. https://archive.epa.gov/emap/archive-emap/web/html/index-158.html.

U.S. Fish and Wildlife Service [USFWS]. 2020. Quality Assurance Project Plan eDNA Monitoring of Bighead and Silver Carps. Page 91.

U.S. Fish and Wildlife Service [USFWS]. 2022, October 6. USFWS National Wetlands Inventory. Geodatabase.

U.S. Geological Survey [USGS]. 1998, August 16. 3D Elevation Program 10-Meter Resolution Digital Elevation Model. Raster, Earth Engine Data Catalog.

U.S. Geological Survey [USGS]. 2021. Watershed Boundary Dataset. Geodatabase.

Valavi, R., J. Elith, J. J. Lahoz-Monfort, and G. Guillera-Arroita. 2021. Modelling species presence-only data with random forests. Ecography 44(12):1731–1742.

Valentini, A., P. Taberlet, C. Miaud, R. Civade, J. Herder, P. F. Thomsen, E. Bellemain, A. Besnard, E. Coissac, F. Boyer, C. Gaboriaud, P. Jean, N. Poulet, N. Roset, G. H. Copp, P. Geniez, D. Pont, C. Argillier, J.-M. Baudoin, T. Peroux, A. J. Crivelli, A. Olivier, M. Acqueberge, M. L. Brun, P. R. Møller, E. Willerslev, and T. Dejean. 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology 25(4):929–942.

Vignali, S., A. G. Barras, R. Arlettaz, and V. Braunisch. 2020. SDMtune: An R package to tune and evaluate species distribution models. Ecology and Evolution 10(20):11488–11506.

Walsh, C. J., A. H. Roy, J. W. Feminella, P. D. Cottingham, P. M. Groffman, and R. P. Morgan. 2005. The urban stream syndrome: current knowledge and the search for a cure. Journal of the North American Benthological Society 24(3):706–723.

Watson, J. M., S. M. Coghlan Jr., J. Zydlewski, D. B. Hayes, and I. A. Kiraly. 2018. Dam Removal and Fish Passage Improvement Influence Fish Assemblages in the Penobscot River, Maine. Transactions of the American Fisheries Society 147(3):525–540.

Whittier, T. R., D. B. Halliwell, and S. G. Paulsen. 1997. Cyprinid distributions in Northeast U.S.A. lakes: evidence of regional-scale minnow biodiversity losses. Canadian Journal of Fisheries and Aquatic Sciences 54:15.

Whittum, K. A., J. D. Zydlewski, S. M. Coghlan Jr, D. B. Hayes, J. Watson, and I. Kiraly. 2023. Fish Assemblages in the Penobscot River: A Decade after Dam Removal. Marine and Coastal Fisheries 15(1):e10227.

Wick, C. 2007. A guide to the freshwater fishes of Maine. Maine Dept. of Inland Fisheries and Wildlife.

Wiken, E., F. J. Nava, and G. E. Griffith. 2011. North American Terrestrial Ecoregions—Level III. Commission for Environmental Cooperation, Montreal, Canada.

Wilcox, T. M., K. S. McKelvey, M. K. Young, S. F. Jane, W. H. Lowe, A. R. Whiteley, and M. K. Schwartz. 2013. Robust detection of rare species using environmental DNA: the importance of primer specificity. PLoS ONE 8(3):e59520.

Wilson, I. G. 1997. Inhibition and facilitation of nucleic acid amplification. Applied and Environmental Microbiology.

Wood, Z. T., B. F. Erdman, G. York, J. G. Trial, and M. T. Kinnison. 2020. Experimental assessment of optimal lotic eDNA sampling and assay multiplexing for a critically endangered fish. Environmental DNA 2(4):407–417.

Wood, Z. T., A. Lacoursière-Roussel, F. LeBlanc, M. Trudel, M. T. Kinnison, C. Garry McBrine, S. A. Pavey, and N. Gagné. 2021. Spatial heterogeneity of eDNA transport improves stream assessment of threatened salmon presence, abundance, and location. Frontiers in Ecology and Evolution 9:1–16.

Yoder, C. O., L. E. Hersha, and E. T. Rankin. 2009. Fish Assemblage and Habitat Assessment of the Presumpscot River. University of Southern Maine, Casco Bay Estuary Partnership, MBI Technical Report MBI/2008-12-6, Portland, ME.

Yoder, C. O., L. E. Hersha, and E. T. Rankin. 2010. The Maine Rivers Fish Assemblage Assessment: Application to the Presumpscot River in 2006 (2010 State of the Bay Presentation).

York, G. 2016, December. Environmental DNA Detection of Invasive Species. Master's Thesis, University of Maine, Orono, ME.

APPENDICES

APPENDIX A: eDNA CORE LAB EXTRACTION STANDARD OPERATING PROCEDURE

A. PURPOSE

The purpose of this standard operating procedure (SOP) is to describe the procedures required to use DNeasy® Blood and Tissue Kit for the extraction of eDNA from filters.

B. SAFETY PRECAUTIONS

- Use standard laboratory PPE
- Use sterilized, filtered pipette tips
- Ensure centrifuge is properly balanced

C. EQUIPMENT AND MATERIAL REQUIRED

1. Laboratory PPE
2. DNeasy® Blood and Tissue Kit
    - Proteinase K
    - Buffer ATL
    - Buffer AL
    - Buffer AW1
    - Buffer AW2
    - Buffer AE
    - 2 ul collection tubes
3. Qiagen Investigator Lyse & Spin Basket kit
4. 1.5 ul microcentrifuge tubes
5. Sterile forceps (1 per filter)
6. Hazardous waste bags/collection vessel
7. Fine tipped marker
8. Timer
9. Pipette(s)
10. Sterile box of filtered pipette tips for each size of pipettes used
11. Shaking incubator
12. Microcentrifuge
13. Vortexer
14. bleach wipes
15. 10% bleach solution
16. Kim Wipes
17. DNA-off
18. UV sterilization light
19. 100% ethanol
20. Gloves, like so many.

D. PROCEDURES

1. Be advised: This SOP will not tell you to change your gloves constantly, but you should. If your glove touches any liquid that contains DNA, change them.
2. Be advised: Incubation times are suggested, but may be changed. They MUST be standard across all samples to be compared. This includes digestion and elution incubations.
3. Clean the work station with Clorox Bleach wipes and DNA-Off and UV sterilization prior to and after all eDNA work.
4. Turn on the incubator and set the temperature to 56°C. Set the shaker setting to low speed. Allow to come up to temperature before continuing.
5. Place the Buffer ATL into the shaker in order to dissolve into solution any precipitate that has formed. This is only necessary if solutes have formed
6. Create an extraction blank
    - A single 47mm glass fiber filter, rolled and placed into a 1.5 ul microcentrifuge tube, as environmental samples.
7. Add 370 ul of Buffer ATL and 30 ul of Proteinase K to each 1.5 ul microcentrifuge tube containing a filtered sample. Vortex immediately and vigorously for 15 seconds. Place samples into a clean tube tray and into shaker incubator.
8. Incubate for 1 hour (timed)
9. Clean workstation with 10%bleach, DNA-off and UV sterilization before continuing.
10. Pre-label all tubes during incubation:
    - 1 spin column + collection tube per filter
    - 1 x 1.5 ul tube for lysis collection
    - 1 lyse & spin basket per filter
    - 1 x 1.5 ul tube for final elution
    - Set aside 3 X 2ul collection tubes per filter (no need to label)
11. Prepare work station with 1 sterile forceps per filter. Have hazardous waste bag/vessel prepared for containing used forceps.
12. Using sterile forceps, press the lysed filter down to expel some of the lysis solution. Using a 1000ul pipette, draw off as much of the lysis as possible, depositing it into a sterile labeled 1.5 ul tube. Seal the tube, this lysis will be used.
13. Using the same forceps, remove the filter and place into the basket (spin column) of the Lyse & Spin basket. If there is any residual lysis, pipette and expel into the 1.5ul tube containing the rest of the lysis.
14. Repeat steps 9+10 for all filters.
15. Spin all Lyse & Spin baskets at maximum speed for 2 minutes. When complete the filters should appear white and dry.
    - If the filters have not given up all the lysis, they may be re-spun. If the second spin isn't sufficient, a new spin basket may be used. All lysis must be retained.
16. Transfer the lysis into the collection tube so all lysis for each filter is within a single tube. Discard filter and basket.
17. Add 200 ul of AL buffer and 200 ul of 100% ethanol to each tube. Vortex immediately for 15 seconds.

18. Transfer 650 ul of the buffer mix to a labeled DNeasy Blood and Tissue spin column. Repeat for all samples.
19. Spin tubes for at 6000xg (8000 rpm) for 1 minute.
20. Discard flow through. Repeat steps 15+16 until all buffer mixture is run through spin column.
21. Add 500ul of AW1 buffer. Spin at 6000xg (8000 rpm) for 1 minute. Discard flow through and place spin column into a fresh collection tube.
22. Add 500ul of AW2 buffer. Spin at maximum speed for 2 minutes. Discard flow through and place spin column into a fresh 2ul collection tube.
23. Spin at maximum speed for 1 minute. Discard flow through carefully, ensure spin column does not contact any flow through at this point. Change to a fresh sterile collection tube.
24. Add 100 ul of AE Buffer. Incubate without spinning at 56°C for 2 minutes.
- Longer incubation times may be used if there is concern for low DNA yield, however times should be standardized across each project.

25. Spin at 6000xg (8000 rpm) for 1 minute.
26. Discard spin column. Transfer flow through from 2 ul collection tube to labeled, sterile 1.5 ul microcentrifuge tubes.
27. Store DNA extract at -20°C for short term use or -80°C for archival or long term storage.
28. Clean entire work station with 10% bleach solution, DNA-off and UV sterilization upon completion. Clean pipettes with DNA-off. Soak tube trays and forceps in 10% bleach solution for a minimum of 10 minutes, then rinse thoroughly with RO/DI water. Forceps should be packed for autoclaving and autoclaved at convenience. Items should finally be UV sterilized in the cabinet sterilizer.

E. QUALITY CONTROL

1. Extraction negative should not amplify in final PCR results

F. NOTES FOR eDNA WORK:

a. Gloves should be changed regularly. At any point if user has touched the inside of a cap or a potentially contaminated area, gloves must be changed. Ensure attention to detail in this area, and be especially cautious of not contaminating reagents and equipment. Filter transfer is the most critical step in terms of cross-contamination between samples and an unclean workstation.
b. Reagents must be treated carefully. Do not re-use tips. Always change tips.
c. Developing a work plan before starting is vital. Set up workspace in a way that is easy and comfortable, but most importantly will decrease the likelihood of contamination. Be prepared before you start.
d. Ensure that any bleach on the bench top is dry before resuming work.
e. Contamination risks at bench:
    i. Air flow. Turn off hood air flow before doing any eDNA work.
    ii. Nothing should be above an open tube or sample. Be aware of how a workstation is set up and avoid a set up that would require moving a hand, sleeve or pipette over an open tube.
    iii. If you have any question, change tips and gloves.

iv. Don’t touch anything. If you do, change your gloves.
v. Change your gloves.

APPENDIX B: PRIMER DEVELOPMENT

As of 2021, no species-specific genetic primer for bridle shiners had been developed. In February 2021 we obtained fifteen bridle shiner fin clips from the New York State Museum for genetic sequencing. Samples were collected in 2015 and 2016 from Cayuga Lake, Sweezy Pond, and Lakeview Pond. We transferred the samples in 95% ethanol and stored them in a -80°C freezer until sequencing.

G. York designed qPCR primers and a Taqman MGB probe to be specific to Bridle Shiner (*Notropis bifrenatus*). *In silico* design was performed using Benchling (Benchling, www.benchling.com) for alignments. We tested against eighteen species *in silico* on the Benchling platform followed by NCBI Primer BLAST. Predicted Tm, primer-dimer and secondary structure potential were tested via IDT OligoAnalyzer (OligoAnalyzer). Multiple mitochondrial gene regions were selected for testing, and our final primers are on cytochrome b and produce an amplicon of 149 base pairs.

In-lab testing included specificity against twenty-one species. Non-target species tissues were identified and collected locally by Maine Inland Fisheries and Wildlife or University of Maine students. Candidate primers were selected for specificity and amplification efficiency. Final primers were optimized for concentration and annealing temperature on Bio-Rad CFX96.

Forward primer 5'-3': TTCACTCCAGCGAACCCC

Reverse primer 5'-3': GGGACTACTAACAGTACTAGGATACTG

Probe 5'-3': GCCACCACACATCCAACCT

Table B.1 Species (n = 18) and accession numbers used for *in silico* testing for the bridle shiner qPCR primer.

Common name	Scientific name	Accession
Blacknose dace	*Rhinichthys atratulus*	AP012104
Blacknose shiner	*Notropis heterolepis*	MG570413

Table B.1 Continued.

Bridle shiner	*Notropis bifrenatus*	MG570408
Bridle shiner	*Notropis bifrenatus*	MG570409
Bridle shiner	*Notropis bifrenatus*	MG570451
Central mudminnow	*Umbra limi*	KP013095
Common shiner	*Luxilus cornutus*	AP012090
Creek chub	*Semotilus atromaculatus*	AP012107
Eastern silvery minnow	*Hybognathus regius*	GQ275151
Fathead minnow	*Pimephales promelas*	KT289925
Finescale dace	*Phoxinus neogaeus*	EU755058
Golden shiner	*Notemigonus crysoleucas*	MG570425
Longnose dace	*Rhinichthys cataractae*	MG570446
Mummichog	*Fundulus heteroclitus*	KT869378
Northern redbelly dace	*Phoxinus eos*	AP009151
Pearl dace	*Margariscus margarita*	AP012081
Rudd	*Scardinius erythrophthalmus*	AP011263
Splendid darter	*Etheostoma barrenense*	AF288424
Spottail shiner	*Notropis hudsonius*	MG570443
Swamp darter	*Etheostoma fusiforme*	FJ937010
Swamp darter	*Etheostoma fusiforme*	HQ128138
Tessellated darter	*Etheostoma olmstedi*	MH301061

Table B.2 Species (n = 21) used for in vitro lab validation testing of the bridle shiner qPCR primer.

Common name	Scientific name
Alewife	*Alosa pseudoharengus*
American eel	*Anguilla rostrata*
American shad	*Alosa sapidissima*
Atlantic salmon	*Salmo salar*
Atlantic sturgeon	*Acipenser oxyrynchus oxyrynchus*
Blacknose shiner	*Notropis heterolepis*
Bridle shiner	*Notropis bifrenatus*
Brook trout	*Salvelinus fontinalis*
Chain pickerel	*Esox niger*
Common goldfish	*Carassius auratus*
Fathead minnow	*Pimephales promelas*
Finescale dace	*Phoxinus neogaeus*
Largemouth bass	*Micropterus salmoides*
Mummichog	*Fundulus heteroclitus*
Pumpkinseed	*Lepomis gibbosus*
Rainbow smelt	*Osmerus mordax*

Table B.2 Continued.

Rainbow trout	*Oncorhynchus mykiss*
Smallmouth bass	*Micropterus dolomieu*
Striped bass	*Morone saxatilis*
Swamp darter	*Etheostoma fusiforme*
Tomcod	*Microgadus tomcod*
White perch	*Morone americana*

APPENDIX C: FISH COMMUNITY COMPOSITION

We analyzed abundance data from seine net surveys using nonmetric multidimensional scaling (NMDS; Faith et al. 1987) with a Bray-Curtis dissimilarity index using *vegan* (Oksanen et al. 2022) in Program R (version 4.2.2; R Core Team 2022). Following Watson et al. (2018) we used fourth root-transformed catch per seine net haul abundance values to reduce the influence of abundant species and accentuate differences in species assemblages (Clarke 1993). We conducted a single ordination, then plotted both site and species scores. The *metaMDS* function in *vegan* uses principal component analysis to rotate the NMDS axes so that Axis 1 reflects the primary sources of variation in the data (Watson et al. 2018; Oksanen et al. 2022). We used the *envfit* function to determine which species were the intrinsic drivers of the site distribution pattern (Grieger 2019).

The NMDS analyses resulted in a stress value less than 0.20 (final stress = 0.079), indicating that the ordination represented the data well in two dimensions (Clarke 1993). We grouped sites with/without bridle shiner presence using the *ordiellipse* function in *vegan* (Figure C.1). Positive values along Axis 1 were associated with stream dwelling species such as blacknose dace (*Rhinichthys atratulus*) and brook trout (*Salvelinus fontinalis*), while negative values were associated with species found in lentic habitats (e.g., bridle shiner and golden shiner). Bridle shiner, chain pickerel, and largemouth bass were the primary species driving negative values along this axis, and common shiner and white sucker had a significant positive influence on NMDS1 values. We could not determine a consistent pattern for the distribution of species along Axis 2, which was primarily positively influenced by ninespine stickleback, fourspine stickleback, and young-of-year minnows. Negative values along NMDS2 were primarily driven by bridle shiner and common shiner.

Figure C.1 Nonmetric multidimensional scaling ordination including a) sites and b) species ordinations. Site codes can be found in Table 1.1.

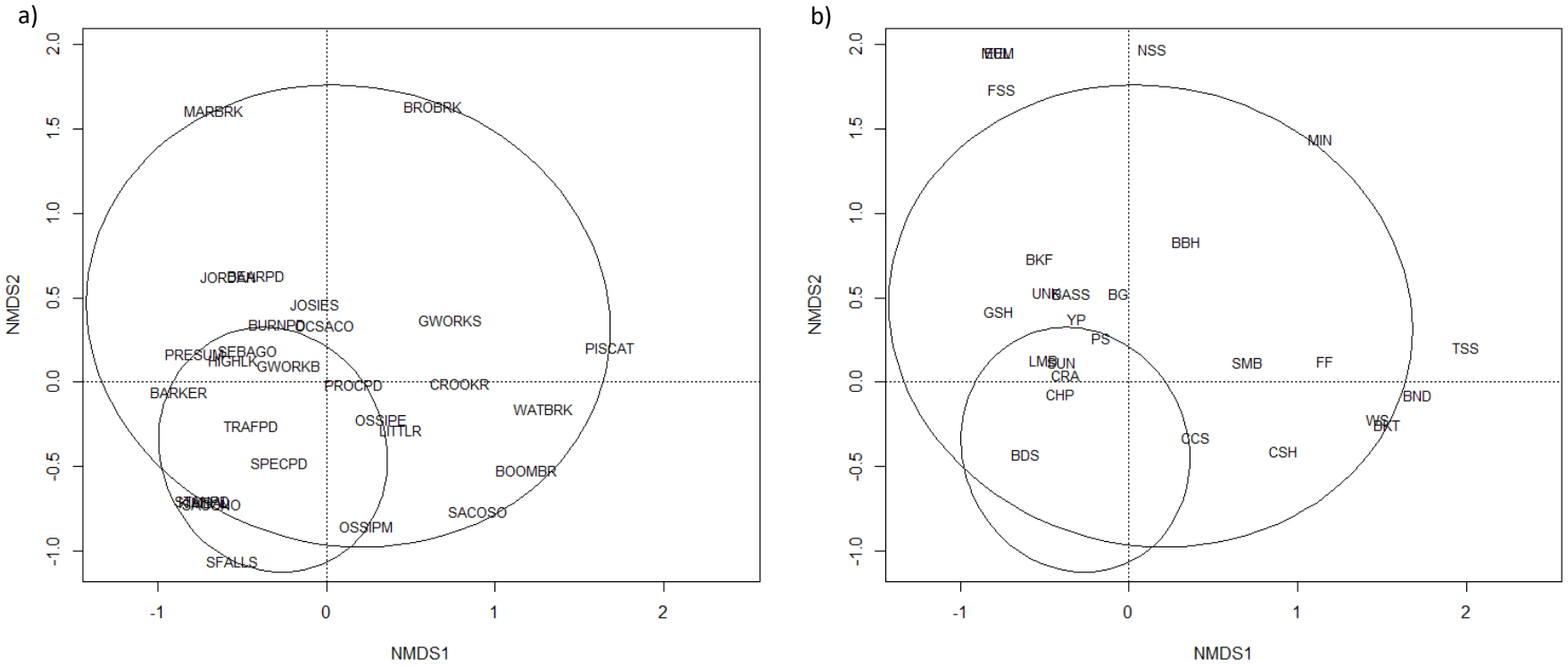
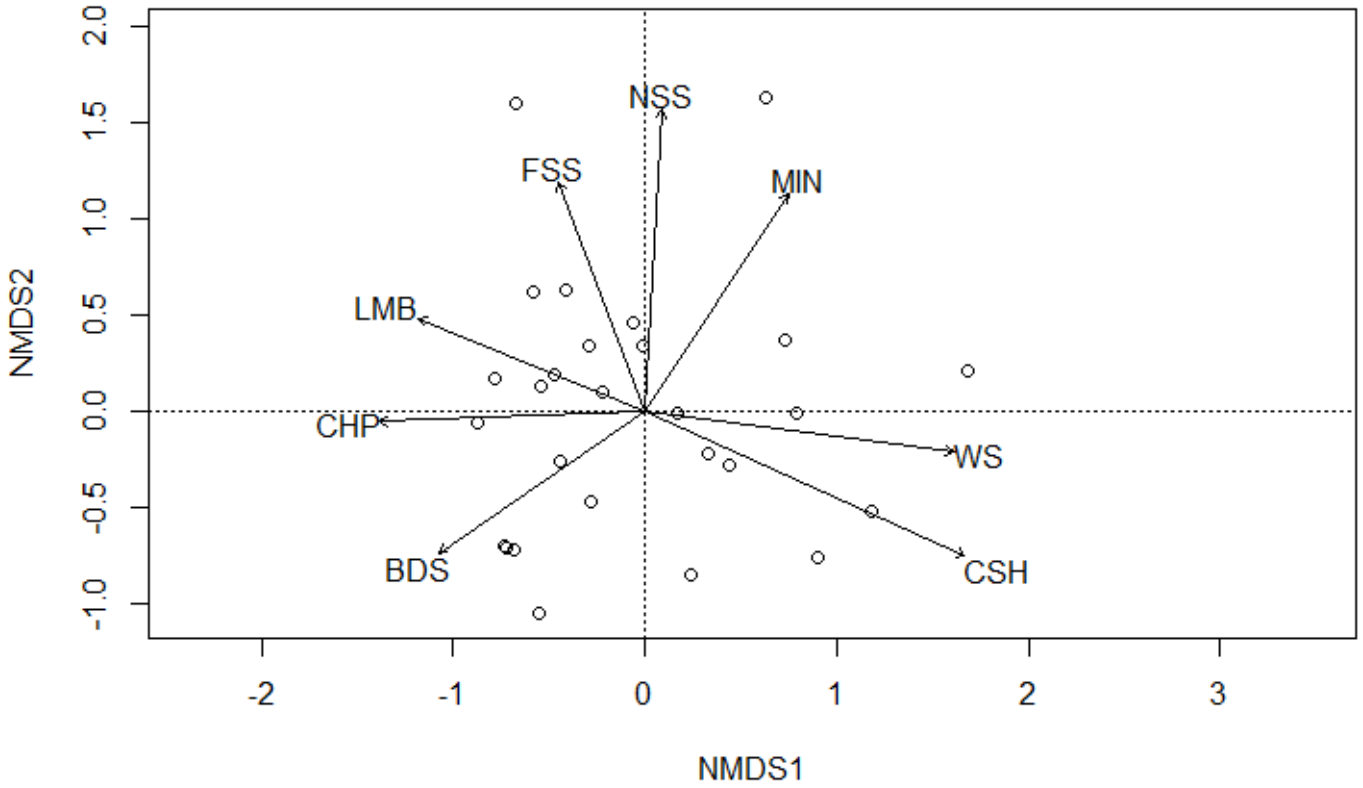


Figure C.2 Species intrinsically driving the site distribution pattern of Figure C.1.



APPENDIX D: HIERARCHICAL OCCUPANCY MODELING CODE

```
#### R code for 3-tiered hierarchical occupancy model in WinBUGS. #####
#### G. York & E. Blomberg, adapted from Kery and Royle 2016
# install.packages("AHMbook")
# install.packages("car")
# install.packages("R2WinBUGS")
# install.packages("plotrix")

# install WinBUGS before continuing

library(AHMbook)
library(car)
library(R2WinBUGS)
library(plotrix)
library(emdbook)
library(ggplot2)
library(gridExtra)

# Import eDNA data ----
pcr1 <- read.csv("./WinBUGS/All_eDNA/BDSoccupancy_pcr1b.csv", header=TRUE,
row.names = 1)
pcr2 <- read.csv("./WinBUGS/All_eDNA/BDSoccupancy_pcr2b.csv", header=TRUE,
row.names = 1)
pcr3 <- read.csv("./WinBUGS/All_eDNA/BDSoccupancy_pcr3b.csv", header=TRUE,
row.names = 1)
pcr4 <- read.csv("./WinBUGS/All_eDNA/BDSoccupancy_pcr4b.csv", header=TRUE,
row.names = 1)

A <- array(as.numeric(NA), dim = c(93,5,4)) #create empty array of 4, 97x5 matrices

A[,,1] <- as.matrix(pcr1) #data into array
A[,,2] <- as.matrix(pcr2)
A[,,3] < -as.matrix(pcr3)
A[,,4] <- as.matrix(pcr4)

y <- A

str( win.data <- list(y = y, # data used to fill array
                      n.site = dim(y)[1], # number of rows = number of sites (97)
                      n.samples = dim(y)[2], # number of columns = number of 1L replicate samples/site
(up to 5)
                      n.pcr = dim(y)[3] )) # number of PCR replicates (4)

# Define model in BUGS language  ----
```

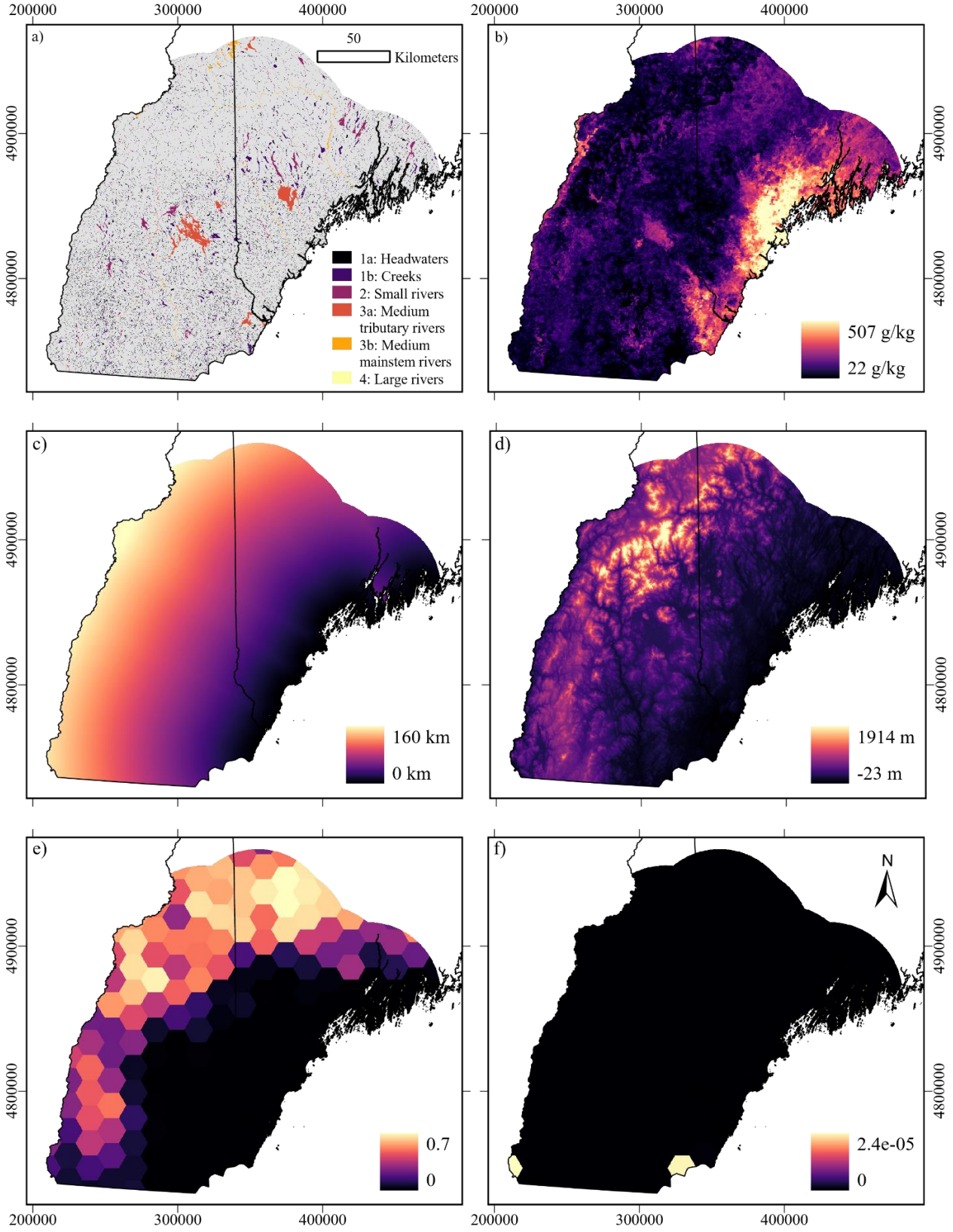
```
sink("eDNA.model.txt")

cat("
model{
# Priors and model for params
  int.psi ~ dunif(0,1) # Intercept of occupancy probability (sites)
  for(t in 1:n.samples){
     int.theta[t] ~ dunif(0,1)} # Intercepts of availability probability per sample replicate

  for(t in 1:n.pcr){
     int.p[t] ~ dunif(0,1)}  # Intercepts of detection probability (1-PCR error)

# 'Likelihood' (or basic model structure)

  ## Occurrence in site i
  for(i in 1:n.site){
     z[i] ~ dbern(psi[i])
     logit(psi[i]) <- logit(int.psi)

     ## Occurrence in sample j
     for(j in 1:n.samples){
        a[i,j] ~ dbern(mu.a[i,j])
        mu.a[i,j] <- z[i] * theta[i,j]
        logit(theta[i,j]) <- logit(int.theta[j])

        ## PCR detection error process in sample k
        for (k in 1:n.pcr){
            y[i,j,k] ~ dbern(mu.y[i,j,k])
            mu.y[i,j,k] <- a[i,j] * p[i,j,k]
            logit(p[i,j,k]) <- logit(int.p[k])
                    }
                  }
             tmp[i] <- step(sum(a[i,])-0.1)
                 }

# Derived quantities
sum.z <- sum(z[]) # Total number of occupied sites
sum.a <- sum(tmp[]) # Total number of samples with presence
mean.p <- mean(int.p[]) # mean p across qPCR replicates
mean.theta <- mean(int.theta[]) # mean theta across sample replicates

} # end model
",fill=TRUE)

sink()
```

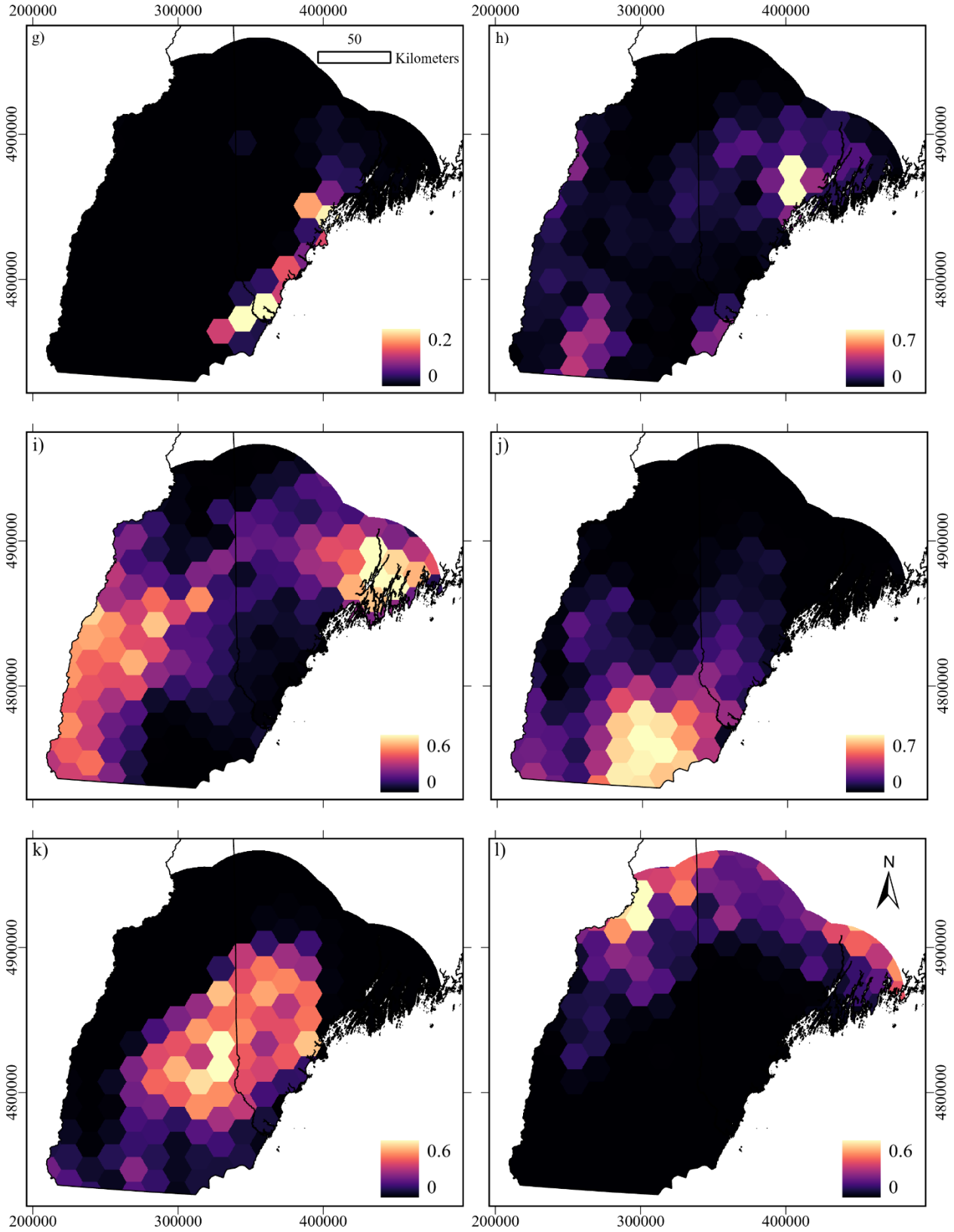
```
# Initial values
zst <- apply(y, 1, max) # inits for presence (z)
ast <- apply(y, c(1,2), max) # inits for availability (a): applies the "max" function to all matrix rows and columns in the array
inits <- function() list(z = zst, a = ast, int.psi = 0.5)

# parameters
params <- c("int.psi", "int.theta", "int.p", "sum.z", "sum.a", "mean.p","mean.theta")

# MCMC setting
ni <- 25000 ; nt <- 10 ; nb <- 2000 ; nc <- 3

# Call WinBUGS and summarize posterior

bd <- "C:/Program Files/winbugs14_full_patched/WinBUGS14" # Location of WinBUGS

eDNA.out <- bugs(win.data, inits, params, "eDNA.model.txt",
                 n.chains = nc, n.thin = nt,
                 n.iter = ni, n.burnin = nb,
                 debug = TRUE, bugs.seed = 42,
                 bugs.dir = bd)

print(eDNA.out, 4)
# Inference for Bugs model at "eDNA.model.txt", fit using WinBUGS,
# 3 chains, each with 25000 iterations (first 2000 discarded), n.thin = 10
# n.sims = 6900 iterations saved
#                  mean      sd     2.5%      25%      50%      75%    97.5%   Rhat  n.eff
# int.psi        0.2010   0.0522   0.1159   0.1653   0.1950   0.2313   0.3181 1.0014  3000
# int.theta[1]   0.5727   0.1318   0.3123   0.4825   0.5747   0.6664   0.8206 1.0011  6900
# int.theta[2]   0.5876   0.1421   0.3086   0.4882   0.5911   0.6890   0.8501 1.0009  6900
# int.theta[3]   0.4010   0.1591   0.1261   0.2840   0.3907   0.5102   0.7256 1.0010  6900
# int.theta[4]   0.6650   0.2355   0.1608   0.4993   0.7028   0.8664   0.9876 1.0014  3900
# int.theta[5]   0.6623   0.2339   0.1619   0.4969   0.6966   0.8594   0.9875 1.0018  2500
# int.p[1]       0.5959   0.0955   0.4040   0.5330   0.5985   0.6613   0.7731 1.0009  6900
# int.p[2]       0.5951   0.0942   0.4086   0.5324   0.5970   0.6622   0.7718 1.0009  6900
# int.p[3]       0.5221   0.0962   0.3368   0.4546   0.5226   0.5897   0.7046 1.0024  1200
# int.p[4]       0.5593   0.0957   0.3691   0.4943   0.5597   0.6280   0.7385 1.0012  4400
# sum.z         18.1671   3.0063  15.0000  16.0000  17.0000  19.0000  26.0000 1.0013  4000
# sum.a         17.8939   2.7686  15.0000  16.0000  17.0000  1 9.0000  25.0000 1.0014  3300
# mean.p         0.5681   0.0514   0.4647   0.5335   0.5692   0.6045   0.6650 1.0015  3000
# mean.theta     0.5777   0.0895   0.3931   0.5171   0.5810   0.6416   0.7407 1.0009  6900
# deviance     139.1212 7.5940 129.8000 132.7000 138.0000 143.7250 157.1525 1.0008  6900
#
# For each parameter, n.eff is a crude measure of effective sample size,
# and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
#
```

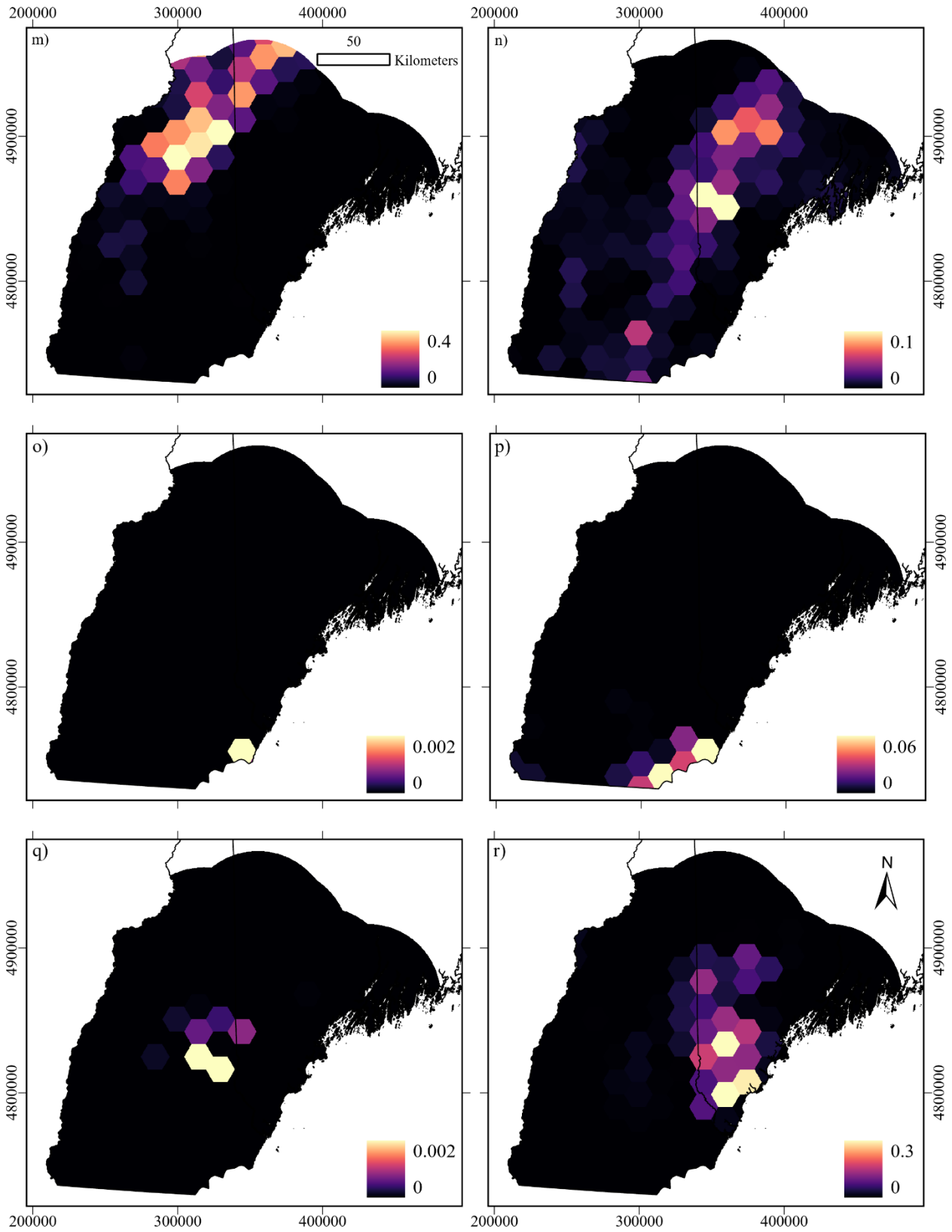
```
# DIC info (using the rule, pD = var(deviance)/2)
# pD = 28.8 and DIC = 168.0
# DIC is an estimate of expected predictive error (lower deviance is better).
#
# # Total # of occupied sites
# 18.1671

# Total # of samples with presence
# 17.8939

# mean probability (psi) of eDNA presence at site
# 0.2010

# mean probability (theta) of collecting eDNA in a single sample
# 0.5777

# mean detection prob. (p) of detecting eDNA in a qPCR rep
# 0.5681

# Calculate p* (Code from E. Blomberg) ----
no.pcr <- seq(1,10,1)

pstar <- 1-((1-eDNA.out$mean$mean.p)^no.pcr)
1-((1-eDNA.out$mean$mean.p)^4)
# p* = 0.9652144 # with 4 PCR replicates
sigma <- matrix(c(eDNA.out$sd$mean.p^2,0,
                  0,0), nrow=2)

pstar.se <- vector(length=length(pstar))

for (i in 1:length(no.pcr)){

pstar.se[i]<- sqrt(deltavar(1-((1-x)^y),
                            meanval = c(x=eDNA.out$mean$mean.p, y=no.pcr[i]),
                            Sigma=sigma)) }

pstar.df <- data.frame(pcr=no.pcr, pstar = pstar, se=pstar.se)

pstar.95 <- 4 # number of PCR replicates needed for detection to exceed 95%
pstar.80 <- 2 # number of PCR replicates needed for detection to exceed 80%
pstar.50 <- 1 # number of PCR replicates needed for detection to exceed 50%

# 95% Confidence Intervals
pstar.df$lower <- pstar.df$pstar - pstar.df$se*1.96
pstar.df$upper <- pstar.df$pstar + pstar.df$se*1.96
pstar.df$upper[pstar.df$upper > 1] <- 1
```

```
write.csv(pstar.df, "./WinBUGS/All_eDNA/pstar.csv")

# Calculate theta* (code from E. Blomberg) ----
no.samples <- seq(1,10,1)

tstar <- 1-((1-eDNA.out$mean$mean.theta)^no.samples)

sigma <- matrix(c(eDNA.out$sd$mean.theta^2,0,
                  0,0), nrow=2)

tstar.se <- vector(length=length(tstar))

for (i in 1:length(no.samples)){
   tstar.se[i]<- sqrt(deltavar(1-((1-x)^y),
                     meanval = c(x=eDNA.out$mean$mean.theta, y=no.samples[i]),
                     Sigma=sigma)) }

tstar.df <- data.frame(samples=no.samples, tstar = tstar, se=tstar.se)

tstar.95 <- 4 # number of 1-L water samples needed for availability to exceed 95%
tstar.80 <- 2 # number of 1-L water samples needed for availability to exceed 80%
tstar.50 <- 1 # number of 1-L water samples needed for availability to exceed 50%

# 95% Confidence Intervals
tstar.df$lower <- tstar.df$tstar - tstar.df$se*1.96
tstar.df$upper <- tstar.df$tstar + tstar.df$se*1.96
tstar.df$upper[tstar.df$upper > 1] <- 1
write.csv(tstar.df, "./WinBUGS/All_eDNA/thetastar.csv")
```

APPENDIX E: SPECIES DISTRIBUTION MODEL PREDICTOR VARIABLES




g)
50
Kilometers
0.2
0
h)
0.7
0
i)
0.6
0
j)
0.7
0
k)
0.6
0
l)
N
0.6
0
200000
300000
400000
4900000
4800000


m)
50
Kilometers
0.4
0
n)
0.1
0
o)
0.002
0
p)
0.06
0
q)
0.002
0
r)
N
0.3
0
200000
300000
400000
4900000
4800000

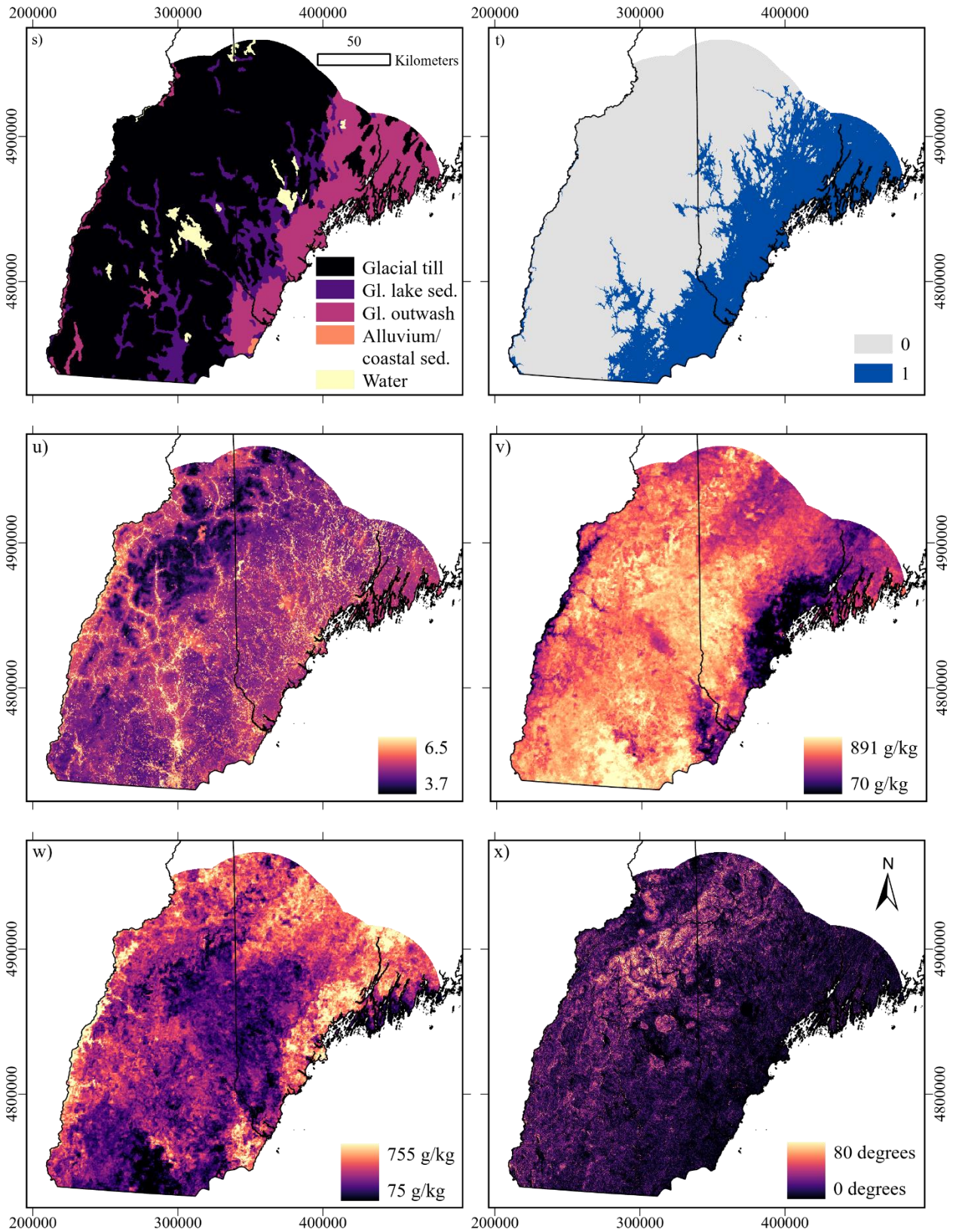


Figure E.1 Predictor variable inputs (n = 24) to bridle shiner species distribution models. Variables were scaled about their means and standard deviations prior to modeling. Maps depict:

a) catchment position, b) soil clay composition (g/kg), c) distance from the coast (km), d) elevation above sea level (meters), e) proportion of Laurentian-Acadian Northern Hardwoods Forest (*for.1920*) in hexagon, f) proportion of Northeastern Interior Dry-Mesic Oak Forest (*for.1921*) in hexagons, g) proportion of Northern Atlantic Coastal Plain Hardwood Forest (for.1922) in hexagon, h) proportion of Laurentian-Acadian Northern Pine(-Oak) Forest in hexagon, i) proportion of Laurentian-Acadian Pine-Hemlock-Hardwood Forest in hexagon, j) proportion of Central Appalachian Dry Oak-Pine Forest in hexagon, k) proportion of Appalachian (Hemlock-)Northern Hardwood Forest in hexagon, l) proportion of Acadian Low-Elevation Spruce-Fir-Hardwood Forest in hexagon, m) proportion of Acadian-Appalachian Montane Spruce-Fir Forest in hexagon, n) proportion of Central Appalachian Pine-Oak Rocky Woodland in hexagon, o) proportion of Northern Atlantic Coastal Plain Maritime Forest in hexagon, p) North-Central Interior Wet Flatwoods in hexagon, q) proportion of Boreal Jack Pine-Black Spruce Forest in hexagon, r) proportion of Northeastern Interior Pine Barrens in hexagon, s) lithology (glacial till coarse, glacial lake sediment fine, glacial outwash coarse, alluvium and coastal sediment fine, water), t) marine limit (above or below 128-m limit), u) soil water pH, v) soil sand content (g/kg), w) soil silt content (g/kg), and x) point slope (degrees).

APPENDIX F: BRIDLE SHINER SURVEY SITES IN MAINE AND NEW HAMPSHIRE

Table F.1 Historic (1898-1999) and current (2000-2022) bridle shiner survey sites (n = 250) in Maine and New Hampshire used as species distribution model inputs. Sites labeled "NA" in the Historic Occupancy or Current Occupancy columns were not included in the model for that time period.

Year Established	Year Last Sampled	State	Waterbody	Site Name	Historic status (1898-1999)	Historic Occupancy	Current status (2000-2022)	Current Occupancy	Point Accuracy	Easting	Northing
1898	Pre-2000	ME	Little Sebago Lake	Little Sebago Lake	Present in waterbody	1	Unknown	NA	Center of waterbody: no precise location known	386517	4859470
1899	2022	ME	Crescent Lake	CRESLK-02	Present in waterbody	1	Present	1	GPS survey point	383820	4869380
1899	Pre-2000	ME	Chaffin Pond	Chaffin Pond	Present in waterbody	1	Unknown	NA	Center of waterbody: no precise location known	384012	4856060
1900	2021	ME	Sebago Lake/ Songo River	SEBAGO-01	Present in waterbody	1	Present	1	GPS survey point	373567	4863550
1937	2021	ME	Crooked River	CROOKR	Present	1	Absent	0	GPS survey point 2021, historic road crossing coordinates	373952	4873110
1937	2021	ME	Great Works River	GWORKN	Present	1	Absent	0	GPS survey point 2021, historic road crossing coordinates	357524	4805680

Table F.1 Continued.

1937	2021	ME	Great Works River	GWORKS	Present	1	Present	1	GPS survey point 2021, historic road crossing coordinates	358935	4797490
1937	2021	ME	Great Works River/ Bauneg Beg Pond	GWORKB	Present	1	Absent	0	GPS survey point	359014	4801930
1937	2021	ME	Josie's Brook	JOSIES	Present	1	Absent	0	GPS survey point 2021, historic road crossing coordinates	369686	4840580
1937	2021	ME	Little River	LITTLR	Present	1	Absent	0	GPS survey point 2021, historic road crossing coordinates	350489	4803790
1937	2021	ME	Salmon Falls River	SFALLS	Present	1	Present	1	GPS survey point	345364	4796430
1938	2021	ME	Bear Pond	BEARPD-01	Present in waterbody	1	Absent	0	GPS survey point, historic point estimated	362593	4890820
1938	2021	ME	Bear Pond	BEARPD-02	Present in waterbody	NA	Absent	0	GPS survey point, historic presence estimated elsewhere in waterbody	363317	4889350
1938	2021	ME	Burnt Meadow Pond	BURNPD-01	Present in waterbody	1	Absent	0	GPS survey point, historic point estimated	348493	4865680

Table F.1 Continued.

1938	2021	ME	Burnt Meadow Pond	BURNPD-02	Present in waterbody	NA	Absent	0	GPS survey point, absent at this point in 1956	348628	4865060
1938	2021	ME	Highland Lake	HIGHLK-02	Present in waterbody	NA	Absent	0	GPS survey point, historic presence estimated elsewhere in waterbody	358466	4884480
1938	2021	ME	Highland Lake	HIGHLK-03	Present in waterbody	NA	Absent	0	GPS survey point, historic presence estimated elsewhere in waterbody	362817	4879480
1938	2021	ME	Highland Lake	HIGHLK-04	Present in waterbody	1	Present	1	GPS survey point	359879	4884760
1938	2021	ME	Sebago Lake	SEBAGO-03	Present in waterbody	NA	Absent	0	GPS survey point, historic presence estimated elsewhere in waterbody	370327	4864980
1938	2021	ME	Sebago Lake	SEBAGO-04	Present in waterbody	NA	Absent	0	GPS survey point, historic presence estimated elsewhere in waterbody	375361	4848550

Table F.1 Continued.

1938	2021	ME	Sebago Lake	SEBAGO-06	Present in waterbody	NA	Absent	0	GPS survey point, historic presence estimated elsewhere in waterbody	381712	4861730
1938	2021	ME	Spectacle Pond	SPECPD	Present	1	Absent	0	GPS survey point	346921	4853590
1938	2021	ME	Stanley Pond	STANPD-01	Present in waterbody	NA	Absent	0	GPS survey point, historic presence estimated elsewhere in waterbody	348533	4854720
1938	2021	ME	Stanley Pond	STANPD-02	Present in waterbody	NA	Absent	0	GPS survey point, historic presence estimated elsewhere in waterbody	348100	4854900
1938	2021	ME	Stanley Pond	STANPD-03	Present in waterbody	1	Absent	0	GPS survey point, historic point estimated	347764	4855830
1938	2021	ME	Trafton Pond	TRAFPD-01	Present in waterbody	NA	Absent	0	GPS survey point, historic presence estimated elsewhere in waterbody	348254	4856280

Table F.1 Continued.

1938	2021	ME	Trafton Pond	TRAFPD-02	Present in waterbody	1	Absent	0	GPS survey point, historic point estimated	347797	4856810
1938	2022	ME	Barker Pond	BARKER	Present	1	Present	1	GPS survey point	359052	4860990
1939	1939	ME	Saco River	SACONO	Present	1	Unknown	NA	Estimated from historic road crossing coordinates, no current survey	352830	4862300
1939	2021	ME	Jordan River	JORDAN	Present	1	Absent	0	GPS survey point 2021, historic road crossing coordinates	382462	4860680
1939	2021	ME	Old Course Saco River	OCSACO	Present	1	Absent	0	GPS survey point 2021, historic road crossing coordinates	346256	4884460
1939	2021	ME	Ossipee River	OSSIPM	Present	1	Present	1	GPS survey point	343737	4850670
1939	2021	ME	Saco River	SACONO-01	Present in vicinity	NA	Absent	0	GPS survey point	353424	4862040
1939	2021	ME	Saco River	SACOSO	Present	1	Absent	0	GPS survey point 2021, historic road crossing coordinates	356722	4851960
1939	2022	ME	Ossipee River	OSSIPE	Present	1	Present	1	GPS survey point	353209	4852140

Table F.1 Continued.

1939	Pre-2000	ME	Ossipee River	OSSIPW	Present	1	Unknown	NA	Estimated from road crossing coordinates, no current survey	340528	4850890
1946	2022	ME	Colcord Pond	COLCPD-01	Presumed present	1	Present	1	GPS survey point	342804	4855520
1947	Unknown	NH	Mill Pond (Oyster River)	Mill Pond	Present in waterbody	1	Absent	0	Center of waterbody: no precise location known	343760	4777090
1947	Unknown	NH	Wheelwright Pond	Wheelwright Pond	Present in waterbody	1	Absent	0	Center of waterbody: no precise location known	336626	4778020
1955	2022	ME	Crescent Lake	CRESLK-03	Present in waterbody	NA	Absent	0	GPS survey point, historic presence estimated elsewhere in waterbody	382678	4871880
1955	2022	ME	Crescent Lake/ Tenny River	CRESLK-01	Present in waterbody	1	Present	1	GPS survey point	382630	4867340
1955	2022	ME	Sokokis Lake	SOKOLK-01	Present in waterbody	NA	Absent	0	GPS survey point, historic presence estimated elsewhere in waterbody	356133	4839970

Table F.1 Continued.

1955	2022	ME	Sokokis Lake	SOKOLK-02	Present in waterbody	NA	Absent	0	GPS survey point, historic presence estimated elsewhere in waterbody	354643	4841400
1955	2022	ME	Sokokis Lake	SOKOLK-03	Present in waterbody	1	Absent	0	GPS survey point, historic point estimated	354682	4841370
1960	2022	ME	Marr Pond	MARRPD-01	Present (likely misidentified)	NA	Absent	0	GPS survey point	476034	4999620
1960	2022	ME	Marr Pond	MARRPD-02	Present (likely misidentified)	NA	Absent	0	GPS survey point	476743	4999440
1992	2021	ME	Marshall Brook	MARBRK	Present; likely introduced	NA	Absent	0	GPS survey point	551554	4902250
1992	2021	ME	Proctor Pond	PROCPD-01	Present in waterbody	NA	Absent	0	GPS survey point, historic presence estimated elsewhere in waterbody	356431	4900500
1992	2021	ME	Proctor Pond	PROCPD-02	Present in waterbody	1	Absent	0	GPS survey point, historic point estimated	356592	4900260

Table F.1 Continued.

1999	1999	ME	Little Pond	Little Pond	Present in waterbody	1	Unknown	NA	Center of waterbody: no precise location known	350801	4884320
2002	2002	ME	Unnamed brook	BLANBR	Present	1	Present	1	GPS survey point	396167	4852400
2002	2021	ME	Piscataqua River	PISCAT	Present	1	Absent	0	GPS survey point 2021, historic road crossing coordinates	394376	4847570
2005	2005	NH	Bixby Pond	20050801-1330-WAP-Little Suncook River-Epsom-SEINE	Present	1	Present	1	GPS survey point	312316	4788270
2005	2005	NH	Cocheco River	20050803-1345-WAP-Cocheco River-Farmington-SEINE	Present	1	Present	1	GPS survey point	331081	4808090
2005	2005	NH	Coffin Brook	20050812-1220-WAP-Coffin Brook-Alton-SEINE	Present	1	Present	1	GPS survey point	319306	4810310
2005	2005	NH	Isinglass River	20051012-1345-WAP-Isinglass River-Barrington-SEINE	Present	1	Present	1	GPS survey point	333231	4789540

Table F.1 Continued.

2005	2005	NH	Jones Brook	20050803-1100-WAP-Jones Brook-Middleton-SEINE	Present	1	Present	1	GPS survey point	334893	4816200
2005	2005	NH	Powwow River	20050707-1200-WAP-Powwow River-South Hampton-SEINE	Present	1	Present	1	GPS survey point	337348	4748990
2005	2005	NH	Soucook River	20050616-1400-WAP-Soucook River-Loudon-SEINE	Present	1	Present	1	GPS survey point	299752	4795540
2005	2005	NH	Suncook River	20050801-1130-WAP-Suncook River-Epsom-SEINE	Present	1	Present	1	GPS survey point	306455	4785960
2005	2005	NH	Suncook River	20050812-1200-WAP-Suncook River-Pembroke-SEINE	Present	1	Present	1	GPS survey point	304380	4781370
2005	2005	NH	Trout Pond	20050919-1515-WAP-Trout Pond-Freedom-SEINE	Present	1	Present	1	GPS survey point	328826	4856330

Table F.1 Continued.

2005	2020	NH	Purity Lake	20200813-1100-BS-Purity Lake-Madison-DIPNET	Present	1	Present	1	GPS survey point	332157	4858600
2006	2006	NH	Exeter River	20060626-1155-WAP-Exeter River-Fremont-SEINE	Present	1	Present	1	GPS survey point	327085	4759240
2006	2006	NH	Isinglass River	20060808-800-NHDES-Isinglass River-Barrington-EFISH	Present	1	Present	1	GPS survey point	337298	4790040
2006	2010	NH	Lamprey River	20100914-1000-BS-Lamprey River-Raymond-DIPNET	Present	1	Present	1	GPS survey point	319633	4768920
2006	2013	NH	Winnipesaukee Lake	20060821-1220-WAP-Lake Winnipesaukee-Moultonborough-SEINE	Present	1	Present; declining	1	GPS survey point	303164	4842570
2006	2019	NH	Soucook River	20190920-900-BS-Soucook River-Loudon-DIPNET	Present	1	Present	1	GPS survey point	299994	4795870

Table F.1 Continued.

2006	2020	NH	Pemigawasset Lake	20200917-1130-BS-Pemigewasset Lake-New Hampton-DIPNET	Present	1	Present	1	GPS survey point	289874	4832800
2006	2021	ME	Presumpscot River	PRESUM-01	Present	1	Present	1	GPS survey point	383533	4845090
2006	2021	ME	Presumpscot River	PRESUM-02	Present	1	Present	1	GPS survey point	383583	4845200
2007	2007	NH	Berrys River	20070628-800-NHDES-Berrys River-Strafford-EFISH	Present	1	Present	1	GPS survey point	332695	4794640
2007	2021	ME	Browns Brook	BROBRK	Present (possibly misidentified)	NA	Absent	0	GPS survey point	554524	4967970
2008	2008	NH	Bonfield Brook	20080717-830-EBTJV-unnamed stream-Madison-EFISH	Present	1	Present	1	GPS survey point	327739	4866770
2008	2008	NH	Soucook River	20080812-1000-WAP-Soucook River-Canterbury-SEINE	Present	1	Present	1	GPS survey point	301052	4806980

Table F.1 Continued.

2009	2009	NH	Black Brook	20090629-1200-IMPOUND-Black Brook-Manchester-EFISH	Present	1	Present	1	GPS survey point	298097	4764890
2009	2009	NH	Winnipesaukee Lake	20090604-1130-BS-Lake Winnipesaukee-Moultonborough-SEINE	Present	1	Present	1	GPS survey point	308008	4844130
2009	2009	NH	Winnipesaukee Lake	20090604-1200-BS-Lake Winnipesaukee-Moultonborough-SEINE	Present	1	Present	1	GPS survey point	308056	4844280
2009	2010	NH	Winnipesaukee Lake	20090604-1300-BS-Lake Winnipesaukee-Moultonborough-SEINE	Present	1	Present	1	GPS survey point	307620	4844580
2009	2010	NH	Winnipesaukee Lake	20090914-1200-WARMWATER-Lake Winnipesaukee-Moultonborough-EBOAT	Present	1	Present	1	GPS survey point	307794	4845010

Table F.1 Continued.

2009	2019	NH	Winnipesaukee Lake	20090604-1100-BS-Lake Winnipesaukee-Moultonborough-SEINE	Present	1	Present	1	GPS survey point	307961	4844050
2009	2019	NH	Winnipesaukee Lake	20090604-1230-BS-Lake Winnipesaukee-Moultonborough-SEINE	Present	1	Present	1	GPS survey point	307541	4843770
2010	2010	NH	Lamprey River	20100914-1200-BS-Lamprey River-Raymond-DIPNET	Present	1	Present	1	GPS survey point	318968	4769400
2010	2010	NH	Lees Pond	20100618-1030-BS-Lees Pond-Moultonborough-DIPNET	Present	1	Present	1	GPS survey point	306392	4845820
2010	2010	NH	Winnipesaukee Lake	20100728-1230-BS-Winnipesaukee Lake-Moultonborough-DIPNET	Present	1	Present	1	GPS survey point	307866	4844680

Table F.1 Continued.

2010	2010	NH	Winnipesaukee Lake	20100729-1030-BS-Winnipesaukee Lake-Moultonborough-DIPNET	Present	1	Present	1	GPS survey point	310247	4838360
2010	2010	NH	Winnipesaukee Lake	20100729-1130-BS-Winnipesaukee Lake-Moultonborough-DIPNET	Present	1	Present	1	GPS survey point	307474	4845060
2010	2019	NH	Winnipesaukee Lake	20100615-1000-BS-Winnipesaukee Lake-Moultonborough-DIPNET	Present	1	Present	1	GPS survey point	305741	4843980
2010	2019	NH	Winnipesaukee Lake	20100615-1100-BS-Winnipesaukee Lake-Moultonborough-DIPNET	Present	1	Present	1	GPS survey point	306109	4844090
2010	2019	NH	Winnipesaukee Lake	20100728-1015-BS-Winnipesaukee Lake-Moultonborough-DIPNET	Present	1	Present	1	GPS survey point	306640	4844450

Table F.1 Continued.

2010	2019	NH	Winnipesaukee Lake	20100728-1100-BS-Winnipesaukee Lake-Moultonborough-DIPNET	Present	1	Present	1	GPS survey point	307118	4844510
2010	2019	NH	Winnipesaukee Lake	20100728-930-BS-Winnipesaukee Lake-Moultonborough-DIPNET	Present	1	Present	1	GPS survey point	306352	4844270
2010	2019	NH	Winnipesaukee Lake	20100728-945-BS-Winnipesaukee Lake-Moultonborough-DIPNET	Present	1	Present	1	GPS survey point	306541	4844270
2010	2019	NH	Winnipesaukee Lake	20190702-1000-BS-Winnipesaukee Lake-Moultonborough-DIPNET	Present	1	Present	1	GPS survey point	306859	4844600
2010	2021	ME	Boom Rd. brook	BOOMBR	Present	1	Absent	0	GPS survey point 2021, historic road crossing coordinates	377144	4819870
2010	2021	ME	Kimball Brook	KIMBAL	Present	1	Present	1	GPS survey point	341689	4887230

Table F.1 Continued.

2010	2022	ME	Watchic Pond brook	WATBRK	Present	1	Absent	0	GPS survey point 2021, historic road crossing coordinates	368055	4845460
2011	2011	NH	Lamprey River	20110811-930-BS-Lamprey River-Raymond-DIPNET	Presumed present	1	Present	1	GPS survey point	324701	4765520
2011	2011	NH	Lamprey River	20110823-1100-BS-Lamprey River-Raymond-DIPNET	Presumed present	1	Present	1	GPS survey point	321419	4767070
2012	2012	NH	Bunker Pond	20120731-1400-BS-Lamprey River-Epping-SEINE	Present	1	Absent	0	GPS survey point	326481	4767360
2012	2012	NH	Jones Brook	20120801-1030-BS-Jones Brook-Middleton-DIPNET	Presumed present	1	Present	1	GPS survey point	333229	4816800
2012	2012	NH	Jones Brook	20120801-1200-BS-Jones Brook-Middleton-DIPNET	Presumed present	1	Present	1	GPS survey point	333751	4817080

Table F.1 Continued.

2012	2012	NH	Jones Brook	20120801-1330-BS-Jones Brook-Middleton-DIPNET	Presumed present	1	Present	1	GPS survey point	332844	4817220
2012	2012	NH	Jones Brook	20120801-900-BS-Jones Brook-Middleton-DIPNET	Presumed present	1	Present	1	GPS survey point	334783	4816370
2012	2012	NH	Lamprey River	20120817-800-BS-Lamprey River-Epping-DIPNET	Presumed present	1	Present	1	GPS survey point	326778	4767710
2012	2019	NH	Branch River	20120802-1300-BS-Branch River-Milton-DIPNET	Presumed present	1	Present	1	GPS survey point	340403	4814010
2013	2013	NH	Crystal Lake	20130718-1000-BS-Crystal Lake-Gilmanton-DIPNET	Presumed present	1	Present	1	GPS survey point	312049	4813770
2013	2013	NH	Garland Pond (west)	20130806-1200-BS-Garland Pond-Moultonborough-DIPNET	Presumed present	1	Present	1	GPS survey point	306200	4846810

Table F.1 Continued.

2013	2013	NH	Isinglass River	20130723-1300-BS-Isinglass River-Barrington-DIPNET	Presumed present	1	Present	1	GPS survey point	332103	4788900
2013	2013	NH	Jones Brook	20130812-1000-BS-Jones Brook-Milton-DIPNET	Presumed present	1	Present	1	GPS survey point	337824	4813440
2013	2013	NH	Jones Brook	20130812-1200-BS-Jones Brook-Milton-DIPNET	Presumed present	1	Present	1	GPS survey point	335218	4815050
2013	2013	NH	Seaver Brook	20130603-1200-EBTJV-Seaver Brook-Plaistow-EFISH	Presumed present	1	Present	1	GPS survey point	329426	4743960
2013	2013	NH	Winnipesaukee Lake	20130509-1100-BS-Winnipesaukee Lake-Moultonborough-DIPNET	Presumed present	1	Present	1	GPS survey point	307548	4844390
2014	2014	NH	Cocheco River	20140908-800-BS-Cocheco River-Farmington-DIPNET	Presumed present	1	Present	1	GPS survey point	330772	4809000

Table F.1 Continued.

2014	2014	NH	Coffin Brook	20140929-1000-BS-Coffin Brook-Alton-DIPNET	Presumed present	1	Present	1	GPS survey point	319744	4809650
2014	2014	NH	Coffin Brook	20140929-1200-BS-Coffin Brook-Alton-DIPNET	Presumed present	1	Present	1	GPS survey point	321001	4810060
2014	2014	NH	Coffin Brook	20140929-1300-BS-Coffin Brook-Alton-DIPNET	Presumed present	1	Present	1	GPS survey point	321740	4810230
2014	2014	NH	Coffin Brook	20140929-800-BS-Coffin Brook-Alton-DIPNET	Presumed present	1	Present	1	GPS survey point	319217	4810690
2014	2014	NH	Exeter River	20140827-1000-BS-Exeter River-Fremont-DIPNET	Presumed present	1	Present	1	GPS survey point	329044	4758860
2014	2014	NH	Exeter River	20140827-800-BS-Exeter River-Fremont-DIPNET	Presumed present	1	Present	1	GPS survey point	326845	4759900

Table F.1 Continued.

2014	2014	NH	Northeast Pond	20140723-1000-BS-Northeast Pond-Milton-DIPNET	Presumed present	1	Present	1	GPS survey point	342338	4810900
2014	2014	NH	Northeast Pond	20140723-800-BS-Northeast Pond-Milton-DIPNET	Presumed present	1	Present	1	GPS survey point	341558	4814410
2014	2014	NH	Ryefield Brook	20140909-1300-BS-Ryefield Brook-Wolfeboro-DIPNET	Presumed present	1	Present	1	GPS survey point	327387	4830760
2014	2014	NH	Warren Brook	20140909-1000-BS-Warren Brook-Wolfeboro-DIPNET	Presumed present	1	Present	1	GPS survey point	328940	4828230
2014	2014	NH	Wentworth Lake	20140909-1200-BS-Wentworth Lake-Wolfeboro-DIPNET	Presumed present	1	Present	1	GPS survey point	327301	4830530
2014	2014	NH	Wentworth Lake	20140909-800-BS-Wentworth Lake-Wolfeboro-DIPNET	Presumed present	1	Present	1	GPS survey point	323743	4828390

Table F.1 Continued.

2015	2015	NH	Heron Pond	20151008-1230-BS-Heron Pond-Moultonborough-DIPNET	Presumed present	1	Present	1	GPS survey point	300604	4845920
2015	2015	NH	Powwow River	20150723-1030-BS-Powwow River-South Hampton-DIPNET	Presumed present	1	Present	1	GPS survey point	337701	4749520
2015	2015	NH	Province Lake	20160830-1230-BS-South River-Effingham-DIPNET	Unknown	NA	Absent	0	GPS survey point	338958	4840540
2015	2015	NH	Unnamed pond	20150924-1015-BS-unnamed pond-Moultonborough-DIPNET	Presumed present	1	Present	1	GPS survey point	299997	4847940
2016	2016	NH	Copp Brook	20160830-930-BS-Copp Brook-Wakefield-DIPNET	Unknown	NA	Absent	0	GPS survey point	338388	4826560
2016	2016	NH	Copps Pond	20160922-1130-BS-Copps Pond-Tuftonboro-DIPNET	Unknown	NA	Absent	0	GPS survey point	316279	4838670

Table F.1 Continued.

2016	2021	NH	Berry Pond	20160922-900-BS-Berry Pond-Moultonborough-DIPNET	Unknown	NA	Absent	0	GPS survey point	307391	4847840
2016	2021	NH	Berry Pond	20210723-1200-BS-Berry Pond-Moultonborough-DIPNET	Unknown	NA	Absent	0	GPS survey point	307193	4848850
2017	2017	NH	Exeter River	20171003-1230-BS-Exeter River-Fremont-DIPNET	Presumed present	1	Present	1	GPS survey point	325972	4760500
2017	2017	NH	Exeter River	20171003-930-BS-Exeter River-Brentwood-DIPNET	Presumed present	1	Present	1	GPS survey point	329771	4759160
2017	2017	NH	Exeter River	20171018-1000-BS-Exeter River-Brentwood-DIPNET	Presumed present	1	Present	1	GPS survey point	336793	4759420
2017	2017	NH	Harper Brook	20170929-930-BS-Harper Brook - New Hampton-DIPNET	Unknown	NA	Absent	0	GPS survey point	287105	4833900

Table F.1 Continued.

2017	2017	NH	Red Hill River	20170926-1030-BS-Red Hill River-Sandwich-DIPNET	Presumed present	1	Present	1	GPS survey point	303589	4851420
2017	2017	NH	Red Hill River	20170926-900-BS-Red Hill River-Sandwich-DIPNET	Presumed present	1	Present	1	GPS survey point	303311	4851840
2017	2017	NH	Salmon Falls River	20170814-1030-BS-Salmon Falls River-Milton-DIPNET	Unknown	NA	Absent	0	GPS survey point	342012	4819280
2017	2017	NH	Suncook River	20170913-1100-BS-Suncook River-Epsom-DIPNET	Unknown	NA	Absent	0	GPS survey point	308338	4790510
2017	2017	NH	Union Meadows Pond	20170814-1230-BS-Union Meadows Pond-Wakefield-DIPNET	Unknown	NA	Absent	0	GPS survey point	336326	4818780
2017	2017	NH	Unnamed pond	20170929-1430-BS-unnamed pond-Meredith-DIPNET	Unknown	NA	Absent	0	GPS survey point	292252	4832670

Table F.1 Continued.

2017	2020	NH	Harper Brook	20200917-900-BS-Harper Brook-New Hampton-DIPNET	Presumed present	1	Present in 2018, not detected in 2020:	0	GPS survey point	289309	4833360
2018	2018	NH	Garland Pond (east)	20190701-1030-BS-Garland Pond-Ossipee-DIPNET	Presumed present	1	Present	1	GPS survey point	323622	4840930
2018	2018	NH	Kanasatka Lake	20180709-1000-BS-Kanasatka Lake-Moultonborough-DIPNET	Presumed present	1	Present	1	GPS survey point	303025	4843210
2018	2018	NH	Kanasatka Lake	20180709-1130-BS-Kanasatka Lake-Moultonborough-DIPNET	Presumed present	1	Present	1	GPS survey point	302156	4844500
2018	2018	NH	Kanasatka Lake	20180709-1300-BS-Kanasatka Lake-Moultonborough-DIPNET	Presumed present	1	Present	1	GPS survey point	301199	4844670
2018	2018	NH	Lamprey River	20180629-930-BS-Lamprey River-Epping-DIPNET	Present	1	Present	1	GPS survey point	326725	4767610

Table F.1 Continued.

2019	2019	NH	Beaver Brook	20190924-1400-BS-Beaver Brook-New Durham-DIPNET	Unknown	NA	Absent	0	GPS survey point	325161	4822400
2019	2019	NH	Northeast Pond	20190809-930-BS-Northeast Pond-Milton-DIPNET	Presumed present	1	Present	1	GPS survey point	341348	4815040
2019	2019	NH	Salmon Falls River	20190809-1000-BS-Salmon Falls River-Milton-DIPNET	Presumed present	1	Present	1	GPS survey point	340682	4815310
2019	2019	NH	Salmon Falls River	20190809-1100-BS-Salmon Falls River-Milton-DIPNET	Presumed present	1	Present	1	GPS survey point	340588	4815330
2019	2019	NH	Salmon Falls River	20190809-1230-BS-Salmon Falls River-Milton-DIPNET	Presumed present	1	Present	1	GPS survey point	340628	4815970
2019	2019	NH	Salmon Falls River	20190809-945-BS-Salmon Falls River-Milton-DIPNET	Presumed present	1	Present	1	GPS survey point	341112	4815260

Table F.1 Continued.

2019	2019	NH	Suncook River	20190920-1100-BS-Suncook River-Barnstead-DIPNET	Presumed present	1	Present	1	GPS survey point	314287	4807800
2019	2019	NH	Unnamed stream	20190701-1130-BS-unnamed stream-Ossipee-DIPNET	Presumed present	1	Present	1	GPS survey point	324576	4840400
2019	2022	NH	Warren Hatchery Pond	20220824-1030-BS-Warren Hatchery Pond-Warren-DIPNET	Absent	0	Introduced	NA	GPS survey point	268205	4866060
2020	2020	NH	Archers Pond	20200813-1300-BS-Archers Pond-Ossipee-DIPNET	Unknown	NA	Absent	0	GPS survey point	327486	4843180
2020	2020	NH	Branch River	20200818-1230-BS-Branch River-Milton-DIPNET	Presumed present	1	Present	1	GPS survey point	339216	4815550
2020	2020	NH	Cooks Pond	20200806-1000-BS-Cooks Pond-Madison-DIPNET	Presumed present	1	Present	1	GPS survey point	326549	4859110

Table F.1 Continued.

2020	2020	NH	Isinglass River	20200707-1000-BS-Isinglass River-Barrington-DIPNET	Presumed present	1	Present	1	GPS survey point	330815	4789640
2020	2020	NH	Isinglass River	20200707-1230-BS-Isinglass River-Barrington-DIPNET	Unknown	NA	Absent	0	GPS survey point	330006	4790140
2020	2020	NH	Isinglass River	20200707-1530-BS-Isinglass River-Barrington-DIPNET	Presumed present	1	Present	1	GPS survey point	330319	4789780
2020	2020	NH	Nippo Brook	20200805-900-BS-Nippo Brook-Barrington-DIPNET	Unknown	NA	Absent	0	GPS survey point	330050	4788390
2020	2020	NH	Pine River	20200729-930-BS-Pine River-Effingham-DIPNET	Unknown	NA	Absent	0	GPS survey point	329392	4846180
2020	2020	NH	Salmon Falls River	20200818-1030-BS-Salmon Falls River-Milton-DIPNET	Presumed present	1	Present	1	GPS survey point	341080	4817410

Table F.1 Continued.

2020	2020	NH	Squam River	20200812-1000-BS-Squam River-Ashland-DIPNET	Unknown	NA	Absent	0	GPS survey point	288770	4843690
2021	2021	NH	Bearcamp Pond	20210723-1000-BS-Bearcamp Pond-Sandwich-DIPNET	Unknown	NA	Absent	0	GPS survey point	308843	4854620
2021	2021	NH	Brindle Pond	20210715-1200-BS-Brindle Pond-Barnstead-DIPNET	Unknown	NA	Absent	0	GPS survey point	318108	4803750
2021	2021	NH	Cawley Pond	20210802-940-BS-Cawley Pond-Sanbornton-DIPNET	Unknown	NA	Absent	0	GPS survey point	289508	4824110
2021	2021	NH	Chocorua Lake	20210712-1000-BS-Chocorua Lake-Tamworth-DIPNET	Unknown	NA	Absent	0	GPS survey point	320804	4862780
2021	2021	NH	Conway Lake	20210716-1100-BS-Conway Lake-Conway-DIPNET	Unknown	NA	Absent	0	GPS survey point	335622	4871630

Table F.1 Continued.

2021	2021	NH	Crystal Lake	20210623-1300-BS-Crystal Lake-Gilmanton - DIPNET	Presumed present	1	Present	1	GPS survey point	313542	4811090
2021	2021	NH	Exeter River	20210721-1100-BS-Exeter River-Exeter-DIPNET	Presumed present	1	Present	1	GPS survey point	341677	4758800
2021	2021	NH	Hawkins Pond	20210722-1200-BS-Hawkins Pond-Center Harbor-DIPNET	Unknown	NA	Absent	0	GPS survey point	293836	4840320
2021	2021	NH	Hermit Lake	20210802-1330-BS-Hermit Lake-Sanbornton-DIPNET	Unknown	NA	Absent	0	GPS survey point	289184	4827440
2021	2021	NH	Horn Pond	20210714-1030-BS-Horn Pond-Wakefield-DIPNET	Unknown	NA	Absent	0	GPS survey point	341178	4825920
2021	2021	NH	Moores Pond	20210712-1200-BS-Moores Pond-Tamworth-DIPNET	Presumed present	1	Present	1	GPS survey point	323202	4858490

Table F.1 Continued.

2021	2021	NH	Purity Lake	20210618-1030-BS-Purity Lake-Eaton-DIPNET	Presumed present	1	Present	1	GPS survey point	332664	4860540
2021	2021	NH	Rocky Pond	20210623-930-BS-Rocky Pond-Gilmanton-DIPNET	Unknown	NA	Absent	0	GPS survey point	301241	4808760
2021	2021	NH	Rollins Pond	20210802-1230-BS-Rollins Pond-Sanbornton-DIPNET	Unknown	NA	Absent	0	GPS survey point	289596	4823150
2021	2021	ME	Saco River	SACONO-02	Present in vicinity	NA	Absent	0	GPS survey point	353642	4862300
2021	2021	NH	Snake River	20210713-1100-BS-Snake River-New Hampton-DIPNET	Presumed present	1	Present	1	GPS survey point	294736	4837720
2021	2021	NH	Soucook River	20210616-1000-BS-Soucook River-Loudon-DIPNET	Presumed present	1	Present	1	GPS survey point	301160	4807380
2021	2021	NH	Upper Suncook Lake	20210715-1000-BS-Upper Suncook Lake-Barnstead-DIPNET	Present	1	Present	1	GPS survey point	314625	4807430

Table F.1 Continued.

2021	2021	NH	White Lake	20210716-1400-BS-White Lake-Tamworth-DIPNET	Unknown	NA	Absent	0	GPS survey point	321375	4855970
2021	2021	NH	Whites Pond	20210715-1400-BS-Whites Pond-Pittsfield-DIPNET	Unknown	NA	Absent	0	GPS survey point	312329	4797020
2021	2021	NH	Wickwas Lake	20210723-1400-BS-Wickwas Lake-Meredith-DIPNET	Presumed present	1	Present	1	GPS survey point	293622	4832450
2021	2021	NH	Winona Lake	20210722-1000-BS-Winona Lake-New Hampton-DIPNET	Unknown	NA	Absent	0	GPS survey point	293446	4838540
2021	2022	ME	Saco River backwater	SACONO-03	Presumed present	1	Present	1	GPS survey point	353523	4862370
2022	2022	ME	Androscoggin River	ANDROS	Unknown	NA	Absent	0	GPS survey point	405085	4877020
2022	2022	ME	Bradley Pond	BRADPD-01	Unknown	NA	Absent	0	GPS survey point	351426	4899890
2022	2022	ME	Bradley Pond	BRADPD-02	Unknown	NA	Absent	0	GPS survey point	351071	4899370
2022	2022	ME	Buck Meadow Brook	BUCKBR	Presumed present	1	Present	1	GPS survey point	350874	4869460
2022	2022	ME	Buff Brook	BUFFBR	Unknown	NA	Absent	0	GPS survey point	356590	4828550
2022	2022	ME	Carsley Brook	CARBRK	Unknown	NA	Absent	0	GPS survey point	368265	4882340

Table F.1 Continued.

2022	2022	ME	Chandler Brook	CHANBR	Unknown	NA	Absent	0	GPS survey point	401963	4862410
2022	2022	ME	Colcord Pond inlet	COLCPD-02	Present in waterbody	NA	Absent	0	GPS survey point	342639	4857680
2022	2022	ME	Crooked River	CROOKN	Unknown	NA	Absent	0	GPS survey point	357338	4900570
2022	2022	ME	Crooked River	CROOKS	Unknown	NA	Absent	0	GPS survey point	374475	4870820
2022	2022	ME	Dingley Brook	DINGLY	Unknown	NA	Absent	0	GPS survey point	378789	4863140
2022	2022	ME	Duck Pond Brook	DUCKIN	Unknown	NA	Absent	0	GPS survey point	358274	4884980
2022	2022	ME	Duck Pond Brook	DUCKNO	Unknown	NA	Absent	0	GPS survey point	357422	4889870
2022	2022	ME	Eddy Brook	EDDYBR	Unknown	NA	Absent	0	GPS survey point	393679	4868000
2022	2022	NH	Exeter River	20220823-1430-BS-Exeter River-Chester-DIPNET	Presumed present	1	Present	1	GPS survey point	320961	4759830
2022	2022	NH	Garland Pond (west)	20220711-1330-BS-Garland Pond-Moultonborough-DIPNET	Presumed present	1	Present	1	GPS survey point	305852	4846900
2022	2022	ME	Great Brook	GRTBRK	Unknown	NA	Absent	0	GPS survey point	346438	4846970
2022	2022	ME	Haley Pond brook	HALEY	Unknown	NA	Unknown	NA	GPS survey point	353738	4844540
2022	2022	ME	Half Moon Pond (Otter Ponds complex)	OTTER-02	Unknown	NA	Absent	0	GPS survey point	378319	4846670

Table F.1 Continued.

2022	2022	ME	Ingalls Pond	INGALS-01	Unknown	NA	Absent	0	GPS survey point	355979	4858140
2022	2022	ME	Ingalls Pond	INGALS-02	Unknown	NA	Absent	0	GPS survey point	356014	4857860
2022	2022	NH	Lamprey River	20220914-1330-BS-Lamprey River-Newmarket-DIPNET	Unknown	NA	Absent	0	GPS survey point	341924	4772630
2022	2022	NH	Long Pond	20220826-1030-BS-Long Pond-Northwood-DIPNET	Unknown	NA	Absent	0	GPS survey point	319112	4790390
2022	2022	ME	Meadow Brook	MEADBR	Unknown	NA	Absent	0	GPS survey point	413718	4869180
2022	2022	ME	Merrill Brook	MERRIL	Unknown	NA	Absent	0	GPS survey point	408744	4855850
2022	2022	ME	Middle Range Pond	RANGE-01	Unknown	NA	Absent	0	GPS survey point	389383	4877000
2022	2022	ME	Middle Range Pond	RANGE-02	Unknown	NA	Absent	0	GPS survey point	388929	4874370
2022	2022	ME	Mosquito Pond	MOSQPD	Unknown	NA	Absent	0	GPS survey point	356524	4906740
2022	2022	ME	Mud Pond	MUDNO	Presumed present	1	Present	1	GPS survey point	358926	4865650
2022	2022	ME	Mud Pond	MUDSO	Unknown	NA	Absent	0	GPS survey point	348459	4830540
2022	2022	ME	Ossipee River	OSSIPR	Unknown	NA	Absent	0	GPS survey point	353408	4852070
2022	2022	ME	Panther Pond	PANTHR-02	Presumed present in waterbody	NA	Absent	0	GPS survey point	381908	4863540

Table F.1 Continued.

2022	2022	ME	Panther Pond	PANTHR-03	Presumed present in waterbody	NA	Absent	0	GPS survey point	383454	4864880
2022	2022	ME	Panther Pond/Tenny River	PANTHR-01	Presumed present	1	Present	1	GPS survey point	382062	4866260
2022	2022	ME	Piscataqua River	PISCDN	Unknown	NA	Absent	0	GPS survey point	395381	4845140
2022	2022	ME	Piscataqua River	PISCUP	Unknown	NA	Absent	0	GPS survey point	394834	4850480
2022	2022	ME	Presumpscot River	PRESBG	Unknown	NA	Absent	0	GPS survey point	383487	4846940
2022	2022	ME	Rachel Carson brook	RACHEL	Unknown	NA	Absent	0	GPS survey point	396681	4827270
2022	2022	ME	Red Brook	REDBRK	Unknown	NA	Absent	0	GPS survey point	391689	4831260
2022	2022	NH	Red Hill River	20220711-1030-BS-Red Hill River-Moultonborough-DIPNET	Presumed present	1	Present	1	GPS survey point	305584	4849760
2022	2022	NH	Red Hill River	20220711-1130-BS-Red Hill River-Moultonborough-DIPNET	Presumed present	1	Present	1	GPS survey point	305648	4849550
2022	2022	ME	Robert's Ridge brook	RIDGEB	Unknown	NA	Absent	0	GPS survey point	363830	4826630
2022	2022	ME	Royal River	ROYAL	Unknown	NA	Absent	0	GPS survey point	398071	4874310
2022	2022	ME	Shepard's River	SHEPR	Unknown	NA	Absent	0	GPS survey point	345344	4866430

Table F.1 Continued.

2022	2022	ME	Snake Pond (Otter Ponds complex)	OTTER-01	Unknown	NA	Absent	0	GPS survey point	378872	4846550
2022	2022	ME	Soper Mill Brook	SOPER	Unknown	NA	Absent	0	GPS survey point	402196	4875470
2022	2022	ME	Symmes Pond	SYMMES-01	Unknown	NA	Absent	0	GPS survey point	348917	4834440
2022	2022	ME	Symmes Pond	SYMMES-02	Unknown	NA	Absent	0	GPS survey point	348864	4834450
2022	2022	ME	The Heath	HEATH-01	Unknown	NA	Absent	0	GPS survey point	382068	4875170
2022	2022	ME	The Heath	HEATH-02	Unknown	NA	Absent	0	GPS survey point	381920	4874680
2022	2022	ME	Unnamed pond	CANCO	Unknown	NA	Absent	0	GPS survey point	396733	4837540
Pre-2000	2020	NH	Heads Pond	20200708-900-BS-Heads Pond-Hooksett-DIPNET	Present	1	Absent	0	GPS survey point	301469	4774900
Pre-2000	Unknown	NH	Canobie Lake	Canobie Lake	Present	1	Absent	0	Center of waterbody: no precise location known	315595	4740760
Pre-2000	Unknown	NH	Lower Suncook Lake	Lower Suncook Lake	Present in waterbody	1	Absent	0	Center of waterbody: no precise location known	315869	4805010
Pre-2000	Unknown	NH	Pleasant Lake	Pleasant Lake	Present in waterbody	1	Absent	0	Center of waterbody: no precise location known	316274	4784430

Table F.1 Continued.

Pre-2000	Unknown	NH	Shadow Lake	Shadow Lake	Present in waterbody	1	Absent	0	Center of waterbody: no precise location known	316955	4743150
Pre-2000	Unknown	NH	Winnipesaukee Lake	Winnipesaukee Lake: Fish Cove	Present	1	Absent	0	Center of cove: no precise coordinates	303333	4835260
Pre-2000	Unknown	NH	Winnisquam Lake	Winnisquam Lake	Present in waterbody	1	Absent	0	Center of waterbody: no precise location known	297610	4823430

APPENDIX G: LOCAL HABITAT MODELS

Table G.1 Top 50 local habitat models ranked by AICc using *glmulti*.

Rank	Variables	AICc	ΔAIC
1	1 + dams + prop.Fveg + Prop.Ag + Prop.Dforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Eveg1	30.602	0.000
2	1 + WBType + TDS + prop.Fveg + Drainage.Area + Prop.Devel + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple + prop.site.Eveg.cat	34.658	4.055
3	1 + WBType + prop.Sveg + prop.Fveg + Drainage.Area + Prop.Ag + Prop.Mforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area	34.803	4.201
4	1 + WBType + TDS + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple + prop.site.Eveg1 + prop.site.Eveg.cat	35.261	4.658
5	1 + WBType + TDS + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple + prop.site.Eveg.cat	35.726	5.123
6	1 + WBType + prop.org.sub + prop.Sveg + prop.Fveg + Drainage.Area + Prop.Ag + Prop.Mforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area	35.765	5.163
7	1 + WBType + TDS + prop.Fveg + Drainage.Area + Prop.Ag + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple + prop.site.Eveg.cat	35.767	5.165
8	1 + WBType + prop.Sveg + prop.Fveg + Drainage.Area + Prop.Ag + Prop.Mforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex	35.900	5.298
9	1 + WBType + prop.Sveg + prop.Eveg + prop.Fveg + Drainage.Area + Prop.Ag + Prop.Mforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area	36.008	5.406
10	1 + WBType + TDS + dams + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple + prop.site.Eveg.cat	36.086	5.484
11	1 + WBType + TDS + prop.Fveg + Drainage.Area + Prop.Wtl + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple + prop.site.Eveg.cat	36.268	5.665
12	1 + WBType + prop.Sveg + prop.Fveg + Drainage.Area + Prop.Ag + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex	36.332	5.730
13	1 + WBType + dams + prop.Sveg + prop.Fveg + Drainage.Area + Prop.Ag + Prop.Mforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area	36.676	6.074
14	1 + WBType + prop.Sveg + prop.Fveg + Drainage.Area + Prop.Ag + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area	36.714	6.111
15	1 + WBType + TDS + prop.Fveg + Drainage.Area + Prop.Devel + Prop.Mforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + Prop.ag.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple	36.736	6.134
16	1 + WBType + TDS + dams + prop.Fveg + Drainage.Area + Prop.Cforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + prop.site.Eveg1 + prop.site.Eveg.broad	36.807	6.205
17	1 + WBType + TDS + prop.Fveg + Drainage.Area + Prop.Mforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple + prop.site.Eveg.cat	37.316	6.714

Table G.1 Continued.

18	1 + WBType + TDS + prop.Sveg + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + prop.site.Eveg.broad	37.325	6.723
19	1 + WBType + TDS + dams + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + prop.site.Eveg1 + prop.site.Eveg.broad	37.419	6.817
20	1 + WBType + TDS + prop.Fveg + Drainage.Area + Prop.Cforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple + prop.site.Eveg.cat	37.481	6.879
21	1 + WBType + TDS + prop.Fveg + Drainage.Area + Prop.Dforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple + prop.site.Eveg.cat	37.554	6.952
22	1 + WBType + TDS + prop.Fveg + Drainage.Area + Canopy + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple + prop.site.Eveg.cat	37.690	7.088
23	1 + WBType + TDS + prop.large.sub + prop.Sveg + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + prop.site.Eveg.broad	37.997	7.394
24	1 + WBType + prop.Sveg + prop.Fveg + Drainage.Area + Prop.Ag + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple	38.030	7.428
25	1 + WBType + TDS + dams + prop.Sveg + prop.Eveg + prop.Fveg + Prop.Devel + Prop.Mforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + prop.site.Eveg1	38.141	7.538
26	1 + WBType + dams + prop.Sveg + prop.Fveg + Drainage.Area + Prop.Ag + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex	38.198	7.596
27	1 + prop.Sveg + prop.Eveg + prop.Fveg + Drainage.Area + Prop.Wtl + Canopy + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Eveg.broad	38.277	7.675
28	1 + WBType + prop.org.sub + prop.Sveg + prop.Fveg + Drainage.Area + Prop.Ag + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex	38.321	7.719
29	1 + WBType + prop.Sveg + prop.Fveg + Drainage.Area + Prop.Ag + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + Prop.ag.HUC12 + HUC12.area + prop.site.Sveg.complex	38.332	7.730
30	1 + WBType + TDS + dams + prop.Fveg + Prop.Mforest + Prop.Wtl + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Eveg.broad	38.375	7.772
31	1 + WBType + prop.Sveg + prop.Fveg + Prop.Ag + Prop.Wtl + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + Prop.ag.HUC12 + HUC12.area + prop.site.Eveg.cat	38.499	7.897
32	1 + prop.large.sub + prop.Sveg + prop.Eveg + prop.Fveg + Drainage.Area + Prop.Wtl + Canopy + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Eveg.broad	38.534	7.932
33	1 + WBType + TDS + dams + prop.Fveg + Drainage.Area + Canopy + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple	38.675	8.073
34	1 + WBType + prop.large.sub + prop.Sveg + prop.Fveg + Drainage.Area + Prop.Ag + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area	38.692	8.090
35	1 + WBType + TDS + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple	38.707	8.104
36	1 + WBType + TDS + prop.Sveg + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + prop.site.Eveg.cat + prop.site.Eveg.broad	38.708	8.106

Table G.1 Continued.

37	1 + WBType + TDS + dams + prop.Eveg + prop.Fveg + Prop.Mforest + Prop.Wtl + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Eveg.broad	38.724	8.122
38	1 + WBType + prop.Sveg + prop.Fveg + Prop.Ag + Prop.Wtl + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + Prop.ag.HUC12 + HUC12.area	38.777	8.174
39	1 + WBType + TDS + dams + prop.large.sub + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Sveg.complex + prop.site.Sveg.simple	38.853	8.251
40	1 + WBType + TDS + dams + prop.Fveg + Drainage.Area + Prop.Dforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + prop.site.Eveg1 + prop.site.Eveg.broad	38.855	8.253
41	1 + WBType + TDS + prop.Sveg + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + prop.site.Sveg.complex + prop.site.Eveg.broad	38.884	8.282
42	1 + WBType + TDS + dams + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + prop.site.Eveg1	38.958	8.356
43	1 + WBType + TDS + prop.Sveg + prop.Fveg + Drainage.Area + Prop.Dforest + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + prop.site.Eveg.broad	38.977	8.375
44	1 + WBType + TDS + prop.org.sub + prop.Sveg + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + prop.site.Eveg.broad	38.982	8.380
45	1 + prop.Sveg + prop.Eveg + prop.Fveg + Drainage.Area + Prop.Cforest + Prop.Wtl + Canopy + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Eveg.broad	39.046	8.444
46	1 + prop.Sveg + prop.Fveg + Drainage.Area + Prop.Wtl + Canopy + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Eveg.cat + prop.site.Eveg.broad	39.131	8.529
47	1 + WBType + TDS + dams + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area + prop.site.Eveg1 + prop.site.Eveg.broad	39.146	8.544
48	1 + WBType + TDS + dams + prop.org.sub + Prop.Cforest + Prop.Mforest + Prop.Wtl + IEI + Prop.dfor.HUC12 + Prop.cfor.HUC12 + Prop.ag.HUC12 + prop.site.Sveg.complex + prop.site.Eveg1	39.294	8.691
49	1 + WBType + TDS + prop.Sveg + prop.Fveg + Drainage.Area + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + prop.site.Sveg.simple + prop.site.Eveg.broad	39.305	8.703
50	1 + WBType + prop.Sveg + prop.Fveg + Prop.Ag + Prop.Cforest + Prop.Wtl + IEI + Prop.mfor.HUC12 + Prop.dfor.HUC12 + Prop.cfor.HUC12 + HUC12.area	39.321	8.719

APPENDIX H: EXPLORATORY GENERALIZED LINEAR MODELS

Table H.1 Top 50 historic period (1898-1999) generalized linear models ranked by AICc using *glmulti*.

Rank	Variables	AICc	ΔAICc
1	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1043.90	0.00
2	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1043.95	0.05
3	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1044.34	0.44
4	1 + catchment + lith + clay + elev + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1044.51	0.61
5	1 + catchment + lith + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1044.53	0.63
6	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1941 + pH + silt + slope	1044.56	0.67
7	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt	1044.64	0.74
8	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1044.81	0.92
9	1 + catchment + lith + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1044.84	0.94
10	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt	1044.94	1.05
11	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1941 + pH + silt + slope	1045.10	1.21
12	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + for.1981 + pH + silt + slope	1045.25	1.35
13	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt	1045.28	1.38
14	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + for.1980 + pH + silt + slope	1045.31	1.41
15	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1930 + for.1941 + pH + silt + slope	1045.34	1.44
16	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + pH + slope	1045.35	1.45
17	1 + catchment + lith + clay + elev + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1045.36	1.47
18	1 + catchment + lith + clay + elev + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt	1045.39	1.49
19	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1941 + for.1980 + pH + silt + slope	1045.42	1.52
20	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt	1045.46	1.56
21	1 + catchment + lith + clay + elev + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1930 + for.1941 + pH + silt + slope	1045.50	1.61
22	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1926 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1045.64	1.75
23	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + for.1981 + pH + silt + slope	1045.66	1.77

Table H.1 Continued.

24	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1930 + for.1941 + pH + silt + slope	1045.69	1.79
25	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1941 + for.1981 + pH + silt + slope	1045.70	1.80
26	1 + catchment + lith + marine + clay + elev + for.1920 + for.1921 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1045.73	1.83
27	1 + catchment + lith + marine + clay + elev + for.1920 + for.1921 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1045.73	1.84
28	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1941 + silt + slope	1045.77	1.88
29	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1926 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1045.89	1.99
30	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + for.1980 + pH + silt + slope	1045.92	2.02
31	1 + catchment + lith + clay + elev + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + for.1980 + pH + silt + slope	1045.93	2.03
32	1 + catchment + lith + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + for.1981 + pH + silt + slope	1045.95	2.06
33	1 + catchment + lith + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1930 + for.1941 + pH + silt + slope	1045.95	2.06
34	1 + catchment + lith + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1926 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1046.08	2.19
35	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1941 + pH + slope	1046.15	2.26
36	1 + catchment + lith + marine + clay + elev + for.1921 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1046.23	2.33
37	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + for.1980 + pH + silt	1046.32	2.42
38	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1925 + for.1926 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1046.34	2.44
39	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1941 + pH + slope	1046.38	2.49
40	1 + catchment + lith + clay + elev + for.1921 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1046.45	2.56
41	1 + catchment + lith + clay + elev + for.1922 + for.1924 + for.1925 + for.1926 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1046.50	2.60
42	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1926 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1046.52	2.63
43	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + silt + slope	1046.53	2.64
44	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1926 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt	1046.71	2.81
45	1 + catchment + lith + marine + clay + elev + for.1920 + for.1921 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt	1046.72	2.82

Table H.1 Continued.

46	1 + catchment + lith + marine + clay + elev + for.1921 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1046.74	2.84
47	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + pH + silt + slope	1046.74	2.85
48	1 + catchment + lith + clay + elev + for.1920 + for.1921 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1941 + pH + silt + slope	1046.76	2.87
49	1 + catchment + lith + clay + elev + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1941 + silt + slope	1046.86	2.96
50	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + pH + silt + slope	1046.86	2.96

Table H.2 Top 50 current period (2000-2022) generalized linear models ranked by AICc using *glmulti*.

Rank	Variables	AICc	ΔAICc
1	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt	849.24	0.00
2	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt	849.44	0.20
3	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt	849.49	0.26
4	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1930 + for.1931 + for.1981 + pH + silt	849.88	0.64
5	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt	849.97	0.73
6	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1930 + for.1931 + for.1981 + pH + silt	850.11	0.87
7	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1931 + for.1981 + pH + silt	850.13	0.89
8	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1930 + for.1931 + for.1981 + pH + silt	850.19	0.96
9	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1930 + for.1931 + for.1980 + for.1981 + pH + silt	850.45	1.21
10	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1926 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt	850.56	1.32
11	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1931 + for.1980 + for.1981 + pH + silt	850.58	1.34
12	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1930 + for.1931 + for.1980 + for.1981 + pH + silt	850.67	1.44
13	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1931 + for.1980 + for.1981 + pH + silt	850.69	1.45
14	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1927 + for.1928 + for.1931 + for.1981 + pH + silt	850.73	1.49
15	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1931 + for.1980 + for.1981 + pH + silt	850.76	1.53
16	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH	850.87	1.64
17	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1930 + for.1931 + for.1981 + pH + silt	850.96	1.72

Table H.2 Continued.

18	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt + slope	851.02	1.78
19	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1930 + for.1931 + for.1981 + pH + silt	851.03	1.80
20	1 + catchment + lith + marine + clay + elev + for.1924 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt	851.05	1.82
21	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt + slope	851.17	1.93
22	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt + slope	851.18	1.95
23	1 + catchment + lith + marine + clay + elev + for.1921 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt	851.24	2.00
24	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1930 + for.1931 + for.1981 + pH	851.25	2.01
25	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1926 + for.1927 + for.1928 + for.1929 + for.1931 + for.1980 + pH + silt	851.31	2.07
26	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1926 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt	851.33	2.09
27	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1926 + for.1927 + for.1928 + for.1929 + for.1931 + pH + silt	851.36	2.12
28	1 + catchment + lith + marine + clay + elev + for.1920 + for.1921 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt	851.42	2.19
29	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1927 + for.1929 + for.1931 + for.1981 + pH + silt	851.45	2.21
30	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1925 + for.1926 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt	851.48	2.24
31	1 + catchment + lith + marine + clay + elev + for.1920 + for.1924 + for.1927 + for.1928 + for.1929 + for.1930 + for.1931 + for.1981 + pH + silt	851.49	2.25
32	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1927 + for.1928 + for.1930 + for.1931 + for.1981 + pH + silt	851.54	2.30
33	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1930 + for.1931 + for.1981 + pH + silt + slope	851.60	2.37
34	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt + slope	851.61	2.37
35	1 + catchment + lith + marine + clay + elev + for.1920 + for.1924 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH + silt	851.64	2.40
36	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1931 + for.1980 + for.1981 + pH	851.72	2.48
37	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1926 + for.1927 + for.1928 + for.1931 + for.1981 + pH + silt	851.77	2.54
38	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1926 + for.1927 + for.1928 + for.1929 + for.1930 + for.1931 + for.1981 + pH + silt	851.81	2.57
39	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1930 + for.1931 + for.1981 + pH + silt + slope	851.82	2.58

Table H.2 Continued.

40	1 + catchment + lith + marine + clay + elev + for.1920 + for.1921 + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1930 + for.1931 + for.1981 + pH + silt	851.84	2.60
41	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1931 + for.1981 + pH	852.01	2.77
42	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1931 + for.1980 + pH + silt	852.01	2.77
43	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1926 + for.1927 + for.1928 + for.1930 + for.1931 + for.1981 + pH + silt	852.03	2.79
44	1 + catchment + lith + marine + clay + elev + for.1920 + for.1921 + for.1922 + for.1924 + for.1927 + for.1928 + for.1930 + for.1931 + for.1981 + pH + silt	852.06	2.82
45	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1926 + for.1927 + for.1928 + for.1929 + for.1930 + for.1931 + for.1980 + pH + silt	852.11	2.87
46	1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1927 + for.1928 + for.1929 + for.1931 + pH + silt	852.30	3.06
47	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1930 + for.1931 + for.1980 + pH + silt	852.35	3.11
48	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1927 + for.1928 + for.1929 + for.1930 + for.1931 + for.1980 + pH + silt	852.38	3.14
49	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1927 + for.1928 + for.1930 + for.1931 + for.1980 + for.1981 + pH + silt + slope	852.38	3.15
50	1 + catchment + lith + marine + clay + elev + for.1920 + for.1922 + for.1924 + for.1925 + for.1926 + for.1927 + for.1928 + for.1929 + for.1931 + for.1980 + for.1981 + pH + silt	852.39	3.15

APPENDIX I: SPECIES DISTRIBUTION MODELING CODE

```
# 1. Standardize rasters
library(raster)

# load and stack rasters ----
files <- list.files(path = "./Predictors/sw_ext/",
                    pattern = "tif",
                    all.files = TRUE,
                    full.names = TRUE,
                    include.dirs = FALSE)

predictors <- raster::stack(files)

names(predictors) <- c("catchment", "clay", "coast", "elev",
                       "for.1920", "for.1921", "for.1922", "for.1924", "for.1925",
                "for.1926", "for.1927", "for.1928", "for.1929", "for.1930",
                 "for.1931", "for.1941", "for.1980", "for.1981", "lith",
                       "marine", "sand", "silt", "slope", "pH")
predictors$pH <- reclassify(predictors$pH, cbind(128, NA)) # Arc exported NAs as 128 for
some reason (and only for this layer)

names <- as.factor(c("catchment", "clay", "coast", "elev",
                       "for.1920", "for.1921", "for.1922", "for.1924", "for.1925",
                "for.1926", "for.1927", "for.1928", "for.1929", "for.1930",
                 "for.1931", "for.1941", "for.1980", "for.1981", "lith",
                       "marine", "sand", "silt", "slope", "pH"))
# Standardize rasters -----
## for continuous variables ----
for (i in names){
  writeRaster(scale(na.omit(predictors[[which(names(predictors) %in% i)]]), center = TRUE),
paste0(paste0("./Predictors/sw_ext/Scaled/", i), "_sw.tif"), overwrite = TRUE)
}

# Load standardized rasters ----
files.st <- list.files(path = "./Predictors/sw_ext/Scaled/",
                       pattern = "tif",
                       all.files = TRUE,
                       full.names = TRUE,
                       include.dirs = FALSE)
```

```
predictors.st <- raster::stack(files.st,
                        as.factor(predictors$lith),
                        as.factor(predictors$marine),
                        as.factor(predictors$catchment))
names(predictors.st)
names(predictors.st) <- c("catchment", "clay", "coast", "elev",
                    "for.1920", "for.1921", "for.1922", "for.1924", "for.1925",
              "for.1926", "for.1927", "for.1928", "for.1929", "for.1930",
               "for.1931", "for.1941", "for.1980", "for.1981", "lith",
                    "marine", "sand", "silt", "slope", "pH")

# Check rasters ----
x11()
par(mfrow=c(1,2))
plot(predictors.st$silt)
plot(predictors.st$sand)
-------------------------------------------------------------------------------------------------------------
---------

# 2. SDMtune exploratory code (example with only current presence-absence data)

#############################################################################
#
  # Current presence-absence data ####

#############################################################################
#
  rm(list = ls()) # remove all objects from R

  # Load libraries ----
  library(dismo)
  library(dplyr)
  library(rasterVis)
  library(rgdal)
  library(rJava)
  library(SDMtune)
  library(sp)
  library(terra)
  library(zeallot)

  # Load scaled rasters into terra ----
```

```
  files <- list.files(path = "./Predictors/sw_ext/Scaled/Comparison/",
                      pattern = "tif",
                      all.files = TRUE,
                      full.names = TRUE,
                      include.dirs = FALSE)
  # Stack rasters in terra ----
  predictors <- terra::rast(files)
  names(predictors) # check order of layers

  names(predictors) <- c("catchment", "clay", "coast",
                         "elev", "for.1920", "for.1921",
                         "for.1922", "for.1924", "for.1925", "for.1926",
                         "for.1927", "for.1928", "for.1929", "for.1930",
                         "for.1931", "for.1941","for.1980", "for.1981", "lith",
                         "marine", "pH", "sand", "silt", "slope")

  # Change categorical variables to factors ----
  predictors$lith <- as.factor(predictors$lith)
  predictors$marine <- as.factor(predictors$marine)
  predictors$catchment <- as.factor(predictors$catchment)
  # Check rasters
  # x11()
  # par(mfrow=c(4,4))
  # plot(predictors[[1:16]])
  # x11()
  # par(mfrow=c(4,4))
  # plot(predictors[[17:32]])
  # x11()
  # par(mfrow=c(2,2))
  # plot(predictors[[33:45]])

  # Read in presence/absence points ----
  bds <- readOGR("./BDSsites_snapped.shp") # snapped to catchment position raster
  bds$cOccu[bds$cOccu == -9999] <- NA # Arc won't let NA's be assigned to numeric columns,
so fix this here
  bds$hOccu[bds$hOccu == -9999] <- NA
  bds$Year_Est[bds$Year_Est == -9999] <- NA
  bds$YearSamp[bds$YearSamp == -9999] <- NA

  # Confirm that all instances of -9999 have been removed
```

```
summary(bds$cOccu) # 7 NA's
summary(bds$hOccu) # 93 NA's, only one confirmed historic absence (other sites unknown)

# Current presence coordinates ----
cPresent <- subset(bds, cOccu == 1) # 122 presences
cPres <- cPresent@coords[,1:2]

# Current absence coordinates ----
cAbsent <- subset(bds, cOccu == 0) # 120 known absences
cAbs <- cAbsent@coords[,1:2]
c.ab <- 10000-(length(cAbsent)-4)

# Generate background coordinates ----
set.seed(42)
# bg_coords <- terra::spatSample(predictors,
#                         size = 50000, # returns 12107 points (many NA's from points on land)
#                         method = "random",
#                         na.rm = TRUE,
#                         xy = TRUE,
#                         values = FALSE)
# saveRDS(bg_coords, "./all_bg_coords_terra.rds") # save for later analyses
bg_coords <- readRDS("./all_bg_coords_terra.rds")
bg_coords_cab <- bg_coords[1:c.ab,] # keep first 10,000 points for Maxent

# Combine background and absence coordinates
c_bg_all <- rbind(bg_coords_cab, cAbs)
# saveRDS(c_bg_all, "./curr_bgab_terra_all.rds") # save points
save.image(file = "cSDMtune_exp_comparison.RData")

# Plot
x11()
plot(predictors$catchment)
points(cAbs)
points(bg_coords_cab)
points(c_bg_all)

# Create SWD object ----
c_data <- prepareSWD(species = "BDS",
                     p = cPres, # 121 presence points
                     a = c_bg_all, # 10004 background & absence points
```

```
                env = predictors,
                categorical = c("catchment", "lith", "marine"))
  # ! 4 locations are NA for some environmental variables and have been discarded
  # Species: BDS
  # Presence locations: 121
  # Absence locations: 10000

  swd2csv(c_data, file_name = "./c_data_exploratory_comparison.csv") # save data as csv

  ## Explore degree of autocorrelation ----
  x11()
  plotCor(c_data,
          method = "pearson",
          cor_th = 0.7)

  # Split data for cross-validation ----
  c(ctrain, cval, ctest) %<-% trainValTest(c_data,
                                           val = 0.2,
                                           test = 0.2,
                                           only_presence = FALSE,
                                           seed = 42) # The only_presence argument is used to split only the
presence and not the background locations (Maxent only)
  save.image(file = "cSDMtune_exp_comparison.RData")

  # Random forest model----
  ## Train a model with default settings ----
  set.seed(42)
  c_default_rf <- train(method = "RF", # can only do classification rf in SDMtune
                        data = ctrain)
  # Species: BDS
  # Presence locations: 73
  # Absence locations: 6000
  # mtry: 6
  # ntree: 500
  # nodesize: 1
  cat("Training auc: ", auc(c_default_rf))
  # Training auc:   # overfitting
  cat("Training TSS: ", tss(c_default_rf))
  # Training TSS:
  cat("Testing auc: ", auc(c_default_rf, test = cval))
```

```
# Testing auc:
cat("Testing tss: ", tss(c_default_rf, test = cval))
# Testing tss:

c_default_rf@model@model$confusion
#      0 1  class.error
# 0 5996 4 0.0006666667
# 1   65 8 0.8904109589
# sensitivity = 15/(15+14) = 0.5172
# specificity = 5986/(5986+58) = 0.9904
# overall accuracy = (15+5986)/(5986+14+58+15) = 98.8
# true skill statistic = sensitivity + specificity - 1 = 0.5076

### Variable importance default ----
write.csv((c_vi_defaultrf <- varImp(c_default_rf,
                            permut = 10)), "./c_vi_default_rfexphexcomparison.csv")

save.image(file = "cSDMtune_exp_comparison.RData")

## K-fold cross-validation ----
c_folds_rf <- randomFolds(ctrain,
                      k = 10,
                      only_presence = FALSE,
                      seed = 42)

c_kfold_rf <- train("RF",
                data = ctrain,
                folds = c_folds_rf)
# Replicates: 4
# Presence locations: 73
# Absence locations: 6000
# mtry: 6
# ntree: 500
# nodesize: 1

saveRDS(c_kfold_rf, "./c_kfold_rfexphexcomparison.rds")

cat("Training AUC: ", auc(c_kfold_rf))
# Training AUC: 1  # overfitting
cat("Testing AUC: ", auc(c_kfold_rf, test = TRUE))
```

```
  # Testing AUC:  0.9288006

  ### Variable importance kfold ----
  write.csv((c_vi_kfoldrf <- varImp(c_kfold_rf,
                                    permut = 10)), "./c_vi_kfold_rfexphexcomparison.csv")

  save.image(file = "cSDMtune_exp_comparison.RData")

  ## Data-driven variable selection 1 ----
  # Remove highly correlated variables
  # SDMtune implements an algorithm that removes highly correlated variables repeating the
following steps:
  #   1. Ranks the variables according to the permutation importance or the percent contribution
(the second method is available only for Maxent models).
  #   2. Checks if the variable ranked as most important is highly correlated with other variables,
according to the given method and correlation threshold. If the algorithm finds correlated
variables it moves to the next step, otherwise checks the other variables in the rank;
  #   3. Performs a leave one out Jackknife test among the correlated variables;
  #   4. Remove the variable that decreases the model performance the least when removed,
according to the given metric on the training dataset.
  set.seed(42)
  c_select_var_rf <- varSel(c_kfold_rf,
                            metric = "auc",
                            test = cval,
                            bg4cor = c_data,
                            method = "pearson",
                            cor_th = 0.7,
                            permut = 10)

  # ✓  The variables clay, elev, for.1931, and silt have been removed
  c_select_var_rf
  # Continuous: coast for.1920 for.1921 for.1922 for.1924 for.1925 for.1926 for.1927 for.1928
for.1929 for.1930 for.1941 for.1980 for.1981 pH sand slope
  # Categorical: catchment lith marine

  saveRDS(c_select_var_rf, "./c_select_var_rfexphexcomparison.rds")

  cat("Training AUC: ", auc(c_select_var_rf))
  # Training AUC:
  cat("Testing AUC after: ", auc(c_select_var_rf, test = cval))
```

```
#> Testing AUC after:

### Variable importance varSel ----
write.csv((varImp(c_select_var_rf, permut = 10)),
"./c_vi_select_var_rfexphexcomparison.csv")
save.image(file = "cSDMtune_exp_comparison.RData")

## Tune hyperparameters-----
getTunableArgs(c_select_var_rf)
# "mtry"     "ntree"    "nodesize"

c_rf <- list(mtry = 1:15,
             ntree = seq(500,2000,200),
             nodesize = 1:15)

c_om_rf <- SDMtune::optimizeModel(c_select_var_rf,
                                  hypers = c_rf,
                                  metric = "auc",
                                  seed = 42)

saveRDS(c_om_rf, "./c_om_rfexphexcomparison.rds")

c_best_model_rf <- c_om_rf@models[[1]]

c_om_rf@results[1, ]
#    mtry ntree nodesize train_AUC  test_AUC   diff_AUC
# 1     9   700        1         1 0.9404747  0.0595253

write.csv(c_om_rf@results, "./c_om_rf_resultsexphexcomparison.csv")

save.image(file = "cSDMtune_exp_comparison.RData")

## Data-driven variable selection 2 ----
set.seed(42)
c_reduced_var_rf <- reduceVar(c_best_model_rf, # cross-validated, tuned model
                              th = 1,     # Contribution threshold (2% permutation importance)
                              metric = "auc", # Metric used to evaluate models
                              test = cval,
                              permut = 10,
```

```
                    use_jk = TRUE) # Use Jackknife AUC test (will not remove variables if it
reduces AUC)

  # ✓ No variables  have been removed

  c_reduced_var_rf
  # Replicates: 10
  # Presence locations: 73
  # Absence locations: 6000
  # mtry: 9
  # ntree: 700
  # nodesize: 1
  # Continuous:  clay elev for.1920 for.1921 for.1922 for.1924 for.1925 for.1926 for.1927
for.1928 for.1929 for.1930 for.1931 for.1980 for.1981 pH slope
  # Categorical: catchment lith marine

  cat("Training AUC: ", auc(c_reduced_var_rf))
  # Training AUC:
  cat("Testing AUC after: ", auc(c_reduced_var_rf, test = cval))
  #> Testing AUC after:

  ### Variable importance reduceVar ----
  write.csv((varImp(c_reduced_var_rf, permut = 10)),
"./c_vi_reduced_var_rfexphexcomparison.csv")

  ## Merge SWD ----
  # Index of the best model
  c_index_rf <- which.max(c_om_rf@results$test_AUC)

  # New train dataset containing only the selected variables
  c_new_train_rf <- c_reduced_var_rf@data

  # Merge data
  c_merged_data_rf <- mergeSWD(c_new_train_rf,
                    cval, # The val dataset contains all the initial environmental variables but the
mergeSWD() function will merge only those that are present in both datasets
                    only_presence = FALSE)

  # Presence locations: 97
  # Absence locations: 8000
```

```
save.image(file = "cSDMtune_exp_comparison.RData")

## Final model ----
set.seed(42)
c_final_rf <- train(method = "RF",
                    data = c_merged_data_rf,
                    mtry = c_om_rf@results[c_index_rf, 1], # mtry = 9
                    ntree = c_om_rf@results[c_index_rf, 2], # ntree = 700
                    nodesize = c_om_rf@results[c_index_rf, 3]) # nodesize = 1
c_final_rf
# Object of class SDMmodel
# Method: RF
# Species: BDS
# Presence locations:  97
# Absence locations:  8000
# Model configurations:
# mtry: 9
# ntree: 700
# nodesize: 1
# Variables:
# Continuous: coast for.1920 for.1921 for.1922 for.1924 for.1925 for.1926 for.1927 for.1928
for.1929 for.1930 for.1941 for.1980 for.1981 pH sand slope
# Categorical: catchment lith marine

save.image(file = "cSDMtune_exp_comparison.RData")

# Evaluate using the held apart testing dataset
cat("Training auc: ", auc(c_final_rf))
# Training auc:  1
cat("Training tss: ", tss(c_final_rf))
# Training tss:  1
cat("Testing auc: ", auc(c_final_rf, test = ctest))
# Testing auc:  0.8805
cat("Testing tss: ", tss(c_final_rf, test = ctest))
# Testing tss:  0.628
caucrf <- SDMtune::auc(c_final_rf, test = ctest) # for weighted model mean
# Testing auc:  0.8805

c_final_rf@model@model$confusion
#      0  1 class.error
```

```
# 0 7983 17   0.0021250
# 1   82 15   0.8453608


### Variable importance final ----
(c_vi_finalrf <- varImp(c_final_rf,
                        permut = 10))
#     Variable Permutation_importance    sd
# 1      coast                 47.1 0.002
# 2  catchment                  21.5 0.001
# 3   for.1927                  17.6 0.001
# 4       sand                  5.6 0.000
# 5      slope                  5.2 0.000
# 6         pH                  2.2 0.000
# 7       lith                 0.6 0.000
# 8   for.1980                    0.1 0.000
# 9   for.1920                    0.0 0.000
# 10  for.1921                    0.0 0.000
# 11  for.1922                    0.0 0.000
# 12  for.1924                    0.0 0.000
# 13  for.1925                    0.0 0.000
# 14  for.1926                    0.0 0.000
# 15  for.1928                    0.0 0.000
# 16  for.1929                    0.0 0.000
# 17  for.1930                    0.0 0.000
# 18  for.1941                    0.0 0.000
# 19  for.1981                    0.0 0.000
# 20    marine                  0.0 0.000
write.csv(c_vi_finalrf, "./c_vi_final_rfexphexcomparison.csv")


# Maxent model----
## Train a model with default settings ----
set.seed(42)
c_default_mx <- train(method = "Maxent",
                      data = ctrain)
# Presence locations: 73
# Absence locations: 6000
# fc: lqph
# reg: 1
```

```
# iter: 500

cat("Training auc: ", auc(c_default_mx))
# Training auc:
cat("Training tss: ", tss(c_default_mx))
# Training tss:
cat("Testing auc: ", auc(c_default_mx, test = cval))
# Testing auc:
cat("Testing tss: ", tss(c_default_mx, test = cval))
# Testing tss:

### Variable importance default ----
(c_vi_defaultmx <- varImp(c_default_mx,
                          permut = 10))

write.csv(c_vi_defaultmx, "./c_vi_default_mxexphexcomparison.csv")
save.image(file = "cSDMtune_exp_comparison.RData")

## K-fold cross-validation ----
c_folds_mx <- randomFolds(ctrain,
                          k = 10,
                          only_presence = FALSE,
                          seed = 42)

c_kfold_mx <- train("Maxent",
                    data = ctrain,
                    folds = c_folds_mx)
# Presence locations: 73
# Absence locations: 6000
# Model configurations:
# fc: lqph
# reg: 1
# iter: 500
saveRDS(c_kfold_mx, "./c_kfold_mxexphexcomparison.rds")

cat("Training AUC: ", auc(c_kfold_mx))
# Training AUC:
cat("Testing AUC: ", auc(c_kfold_mx, test = TRUE))
# Testing AUC:
```

```
### Variable importance kfold ----
(c_vi_kfoldmx <- varImp(c_kfold_mx,
                        permut = 10))

write.csv(c_vi_kfoldmx, "./c_vi_kfold_mxexphexcomparison.csv")

save.image(file = "cSDMtune_exp_comparison.RData")

## Data-driven variable selection 1 ----
# Remove highly correlated variables
set.seed(42)
c_select_var_mx <- varSel(c_kfold_mx,
                          metric = "auc",
                          test = cval,
                          bg4cor = c_data,
                          method = "pearson",
                          cor_th = 0.7)

# ✓ The variables coast, for.1941, and sand have been removed
# Variables:
# Continuous:  clay elev for.1920 for.1921 for.1922 for.1924 for.1925 for.1926 for.1927
for.1928 for.1929 for.1930 for.1941 for.1980 for.1981 pH silt slope
# Categorical: catchment lith marine
save.image(file = "cSDMtune_exp_comparison.RData")

cat("Training AUC: ", auc(c_select_var_mx))
# Training AUC:
cat("Testing AUC: ", auc(c_select_var_mx, test = TRUE))
# Testing AUC:

### Variable importance varSel ----
(c_vi_select_var_mx <- varImp(c_select_var_mx,
                              permut = 10))

write.csv(c_vi_select_var_mx, "./c_vi_select_var_mxexphexcomparison.csv")

## Tune hyperparameters-----
getTunableArgs(c_kfold_mx)
# [1] "fc"   "reg"  "iter"
```

```
c_mx <- list(fc = c("l", "lq", "lh", "lqp", "lqph", "lqpht"),
             iter = seq(300,1100,200),
             reg = seq(0.2,1,0.1))

c_om_mx <- optimizeModel(c_select_var_mx,
                         hypers = c_mx,
                         metric = "auc",
                         seed = 42)

c_best_model_mx <- c_om_mx@models[[1]]

c_om_mx@results[1, ]
#      fc reg iter train_AUC  test_AUC   diff_AUC
# 1

save.image(file = "cSDMtune_exp_comparison.RData")
write.csv(c_om_mx@results, "./c_om_mx_resultsexphexcomparison.csv")

## Data-driven variable selection 2 ----
set.seed(42)
c_reduced_var_mx <- reduceVar(c_best_model_mx,
                              th = 1,
                              metric = "auc",
                              test = cval,
                              permut = 10,
                              use_jk = TRUE)

# ✓ The variables for.1931, marine, for.1920, for.1926, for.1922, for.1921, for.1928, for.1981,
for.1925, and for.1929 have been removed

c_reduced_var_mx
# Model configurations:
# fc:
# reg:
# iter:
# Variables:
# Continuous:
# Categorical: catchment lith

auc(c_reduced_var_mx)
```

```
  #> Training AUC:
  cat("Testing AUC after: ", auc(c_reduced_var_mx, test = cval))
  #> Testing AUC after:

  ### Variable importance reduceVar ----
  write.csv((varImp(c_reduced_var_mx, permut = 10)),
"./c_vi_reduced_var_mxexphexcomparison.csv")

  save.image(file = "cSDMtune_exp_comparison.RData")

  ## Merge SWD ----
  # Index of the best model
  c_index_mx <- which.max(c_om_mx@results$test_AUC)

  # New train dataset containing only the selected variables
  c_new_train_mx <- c_reduced_var_mx@data

  # Merge only presence data
  c_merged_data_mx <- mergeSWD(c_new_train_mx,
                       cval, # The val dataset contains all the initial environmental variables but the
mergeSWD() function will merge only those that are present in both datasets
                       only_presence = FALSE)
  # Presence locations: 97
  # Absence locations: 8000

  ## Final model ----
  set.seed(42)
  c_final_mx <- train("Maxent",
                  data = c_merged_data_mx,
                  fc = c_om_mx@results[c_index_mx, 1], # fc = lqpht
                  reg = c_om_mx@results[c_index_mx, 2], # reg = 0.3
                  iter = c_om_mx@results[c_index_mx, 3]) # iter = 300
  # Presence locations:  97
  # Absence locations:  8000
  # Model configurations:
  #  fc: lqpht
  #  reg: 0.3
  #  iter: 300
  # Variables:
  # Continuous:  clay elev for.1924 for.1927 for.1930 for.1980 pH silt slope
```

```
 # Categorical: catchment lith

 # evaluate using the held apart testing dataset
 auc(c_final_mx)
 # 0.9622481
 cat("Training tss: ", tss(c_final_mx))
 # Training tss:  0.7929201
 caucmx <- auc(c_final_mx, test = ctest)
 # 0.92053
 x11()
 plotROC(c_final_mx, test = ctest)

 cat("Testing tss: ", tss(c_final_mx, test = ctest))
 # Testing tss:  0.7265

 (c_vi_finalmx <- varImp(c_final_mx,
                         permut = 10))
 #     Variable Permutation_importance    sd
 # 1   for.1927                 47.7 0.013
 # 2       elev                12.2 0.007
 # 3      slope                 9.9 0.003
 # 4  catchment                  7.4 0.004
 # 5   for.1930                  5.7 0.005
 # 6   for.1924                  5.0 0.005
 # 7       clay                 3.5 0.001
 # 8       silt                3.1 0.002
 # 9   for.1980                  2.9 0.002
 # 10      lith                 2.2 0.002
 # 11        pH                  0.5 0.002

 ### Variable importance final ----
 write.csv(c_vi_finalmx, "./c_vi_final_mxexphexcomparison.csv")
 write.csv(c_final_mx@model@lambdas, "./c_exploratory_mx_lambdas_comparison.csv")
```

---

3. Exploratory GLM of current presence-absence-background data

```
#############################################################################
# Current presence-absence-background data#####
```

```
#############################################################################
rm(list = ls())

# Load libraries ----
library(BiodiversityR)
library(caret)
library(dismo)
library(dplyr)
library(glmulti)
library(raster)
library(rgdal)
library(tidyverse)
library(vip)

# Load scaled rasters into R ----
files <- list.files(path = "./Predictors/sw_ext/Scaled/Comparison/",
                    pattern = "tif",
                    all.files = TRUE,
                    full.names = TRUE,
                    include.dirs = FALSE)

# Stack rasters in raster ----
predictors <- raster::stack(files)
names(predictors) # check order of layers
names(predictors) <- c("catchment", "clay", "coast",
                       "elev", "for.1920", "for.1921",
                       "for.1922", "for.1924", "for.1925", "for.1926",
                       "for.1927", "for.1928", "for.1929", "for.1930",
                       "for.1931", "for.1941","for.1980", "for.1981", "lith",
                       "marine", "pH", "sand", "silt", "slope")

# Change categorical variables to factors ----
predictors$lith <- as.factor(predictors$lith)
predictors$marine <- as.factor(predictors$marine)
predictors$catchment <- as.factor(predictors$catchment)

# Check rasters
# x11()
# par(mfrow=c(4,4))
# plot(predictors[[1:16]])
# x11()
# par(mfrow=c(4,4))
# plot(predictors[[17:32]])
# x11()
# par(mfrow=c(2,2))
# plot(predictors[[33:36]])
```

```
# Read in presence/absence points ----
bds <- readOGR("./BDSsites_snapped.shp") # snapped to catchment position raster
bds$cOccu[bds$cOccu == -9999] <- NA # Arc won't let NA's be assigned to numeric columns, so fix this here
bds$hOccu[bds$hOccu == -9999] <- NA
bds$Year_Est[bds$Year_Est == -9999] <- NA
bds$YearSamp[bds$YearSamp == -9999] <- NA

# Confirm that all instances of -9999 have been removed
summary(bds$cOccu) # 7 NA's
summary(bds$hOccu) # 93 NA's, only one confirmed historic absence (other sites unknown)
summary(bds$Year_Est)
#  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
#  1898    2005    2014    2001    2021    2022       6
summary(bds$YearSamp) # 0 refers to unknown sampling year, NA refers to sites not sampled after 2000
#   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
#      0    2014    2021    2008    2022    2022       8
# write.csv(bds, "./bdssites_snapped.csv")

# Current presence coordinates ----
cPresent <- subset(bds, cOccu == 1) # 122 known presences
cPres <- cPresent@coords[,1:2]

# Current absence coordinates ----
cAbsent <- subset(bds, cOccu == 0) # 121 known absences, 117 in model extent
cAbsent <- subset(cAbsent, Site_Name != "MARRPD-01") # Remove absences outside of model extent
cAbsent <- subset(cAbsent, Site_Name != "MARRPD-02")
cAbsent <- subset(cAbsent, Site_Name != "MARBRK")
cAbsent <- subset(cAbsent, Site_Name != "BROBRK")
cAbs <- cAbsent@coords[,1:2]
c.ab <- 10000-length(cAbsent)

# Read in background coordinates ----
bg_coords <- readRDS("./all_bg_coords_terra.rds") # use same background points as with MaxEnt and RF
bg_coords_cab <- bg_coords[1:c.ab,] # keep first 10,000 points

# Combine background and absence coordinates ----
c_bg_all <- rbind(bg_coords_cab, cAbs)

# Default GLM ----
## Partition presence and absence data ----
folds <- 5 # percentage split (80/20) as in other models
```

```
{set.seed(42)
  kfold_pres <- kfold(cPres, folds)
  kfold_back <- kfold(c_bg_all, folds)

  # Create training data
  cPres_train <- cPres[kfold_pres != 1, ]
  cBackg_train <- c_bg_all[kfold_back != 1, ]
  pab_train <- c(rep(1, nrow(cPres_train)), rep(0, nrow(cBackg_train)))
  envtrain <- raster::extract(predictors, rbind(cPres_train,cBackg_train))
  envtrain <- data.frame(cbind(pab=pab_train, envtrain))
  envtrain <- transform(envtrain,
                        catchment = as.factor(envtrain$catchment),
                        lith = as.factor(envtrain$lith),
                        marine = as.factor(envtrain$marine))
  # Create testing data
  cPres_test <- cPres[kfold_pres == 1, ]
  prestest <- data.frame(raster::extract(predictors, cPres_test))
  prestest <- transform(prestest,
                        catchment = as.factor(prestest$catchment),
                        lith = as.factor(prestest$lith),
                        marine = as.factor(prestest$marine))
  cBackg_test <- c_bg_all[kfold_back == 1, ]
  abstest <- data.frame(raster::extract(predictors, cBackg_test))
  abstest <- transform(abstest,
                       catchment = as.factor(abstest$catchment),
                       lith = as.factor(abstest$lith),
                       marine = as.factor(abstest$marine))
  # GLM with all variables
  c_default_glm <- glm(pab ~ ., family = binomial(link = "logit"),
                       data = envtrain)}
# Warning message:
#   glm.fit: fitted probabilities numerically 0 or 1 occurred

# Test AUC
c_default_e <- evaluate(p=prestest,a=abstest, c_default_glm)
(default_auc <- slot(c_default_e, "auc"))
# 0.9130417

## Variable importance ----
c_vi_default_glm <- vi_model(c_default_glm, type = "stat") # Z-statistic
write.csv(c_vi_default_glm, "./c_vi_default_glmexpcomparison.csv")

save.image("./cGLM_exploratory_comparison.RData")

# K-fold cross-validation  ----
```

```
# Set random seeds
set.seed(42)
(seeds <- sample.int(1000, 10))
# [1] 561 997 321 153  74 228 146 634  49 128

# Set number of folds
folds <- 5 # percentage split (80% training /20% testing) as in other models

# Create empty outputs to hold results
c_kfold_glm <- list()
c_kfold_e <- list()
kfold_auc <- matrix(NA, nrow = 1, ncol = 1)
coutput <- data.frame(matrix(NA, nrow = 10, ncol = 2)) # Create an empty data.frame
colnames(coutput) <- c("seed", "testAUC")

# Cross-validation
for (i in seq_along(seeds)){
  set.seed(i) # set random seed
  kfold_pres <- kfold(cPres, folds) # partition presence and background data according to the
random seed
  kfold_back <- kfold(c_bg_all, folds)

  # Create training data with 4/5 folds
  cPres_train <- cPres[kfold_pres != 1, ]
  cBackg_train <- c_bg_all[kfold_back != 1, ]
  pab_train <- c(rep(1, nrow(cPres_train)), rep(0, nrow(cBackg_train)))
  envtrain <- raster::extract(predictors, rbind(cPres_train,cBackg_train))
  envtrain <- data.frame(cbind(pab=pab_train, envtrain))
  envtrain <- transform(envtrain,
                        catchment = as.factor(envtrain$catchment),
                        lith = as.factor(envtrain$lith),
                        marine = as.factor(envtrain$marine))

  # Create testing data with 1/5 folds
  cPres_test <- cPres[kfold_pres == 1, ]
  prestest <- data.frame(raster::extract(predictors, cPres_test))
  prestest <- transform(prestest,
                        catchment = as.factor(prestest$catchment),
                        lith = as.factor(prestest$lith),
                        marine = as.factor(prestest$marine))
  cBackg_test <- c_bg_all[kfold_back == 1, ]
  abstest <- data.frame(raster::extract(predictors, cBackg_test))
  abstest <- transform(abstest,
                       catchment = as.factor(abstest$catchment),
                       lith = as.factor(abstest$lith),
                       marine = as.factor(abstest$marine))
```

```
  # GLM with all variables
  c_kfold_glm[[i]] <- glm(pab ~ ., family = binomial(link = "logit"),
                          data = envtrain)

  # Test AUC
  c_kfold_e[[i]] <- evaluate(p=prestest,a=abstest, c_default_glm)
  kfold_auc[i] <- slot(c_kfold_e[[i]], "auc")

  # Populate output data frame
  coutput[i, 1] <- seeds[i]
  coutput[i, 2] <- kfold_auc[i]
}
# Warning messages:
#   1: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 2: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 3: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 4: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 5: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 6: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 7: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 8: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 9: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 10: glm.fit: fitted probabilities numerically 0 or 1 occurred
saveRDS(coutput, "./c_kfold_glm_outputexpcomparison.RDS")

## calculate the mean AUC ----
(aucGLM <- sapply(c_kfold_e, function(x){slot(x, 'auc')}))
# [1] 0.9092500 0.8828125 0.9111458 0.9161875 0.8912500 0.9287292 0.8895625 0.8875833
0.9097708 0.8750833
(mean(aucGLM))
# [1] 0.9001375

## Variable importance ----
c_vi_kfold_glm <- vi_model(c_kfold_glm[[1]], type = "stat") # Z-statistic

c_vi_kfold_glm <- c_vi_kfold_glm[order(c_vi_kfold_glm$Variable),] %>% select(-Sign)

for (i in 2:length(c_kfold_glm)) {
  vi <- vi_model(c_kfold_glm[[i]], type = "stat")
  c_vi_kfold_glm <- c_vi_kfold_glm %>% left_join(., vi, by = "Variable") %>% select(-Sign)
}
# average ranks together
c_vi_kfold_glm$Avg <- rowMeans(c_vi_kfold_glm[, 2:11])
c_vi_kfold_glm <- c_vi_kfold_glm[order(-c_vi_kfold_glm$Avg),]
write.csv(c_vi_kfold_glm, "./c_vi_kfold_glmexpcomparison.csv")
```

```
save.image("./cGLM_exploratory_comparison.RData")

# Remove highly correlated variables (equivalent of varSel) ----
# Keep: catchment, clay, elev, for.1920, for.1921, for.1922, for.1924, for.1925, for.1926,
for.1927, for.1928, for.1929, for.1930, for.1931, for.1980, for.1981, lith, marine, pH, silt, slope
# Remove: coast, for.1941, sand

# Drop layers 1 ----
pred_drop <- dropLayer(predictors, c("coast", "for.1941", "sand"))

# GLM with selected variables (no kfold because glmulti can only accept one model as input)----
{set.seed(42)
  kfold_pres <- kfold(cPres, folds)
  kfold_back <- kfold(c_bg_all, folds)

  # Create training data with selected variables
  cPres_train <- cPres[kfold_pres != 1, ]
  cBackg_train <- c_bg_all[kfold_back != 1, ]
  pab_train <- c(rep(1, nrow(cPres_train)), rep(0, nrow(cBackg_train)))
  envtrain <- raster::extract(pred_drop, rbind(cPres_train,cBackg_train))
  envtrain <- data.frame(cbind(pab=pab_train, envtrain))
  envtrain <- transform(envtrain,
                        catchment = as.factor(envtrain$catchment),
                        lith = as.factor(envtrain$lith),
                        marine = as.factor(envtrain$marine))
  # Create testing data
  cPres_test <- cPres[kfold_pres == 1, ]
  prestest <- data.frame(raster::extract(pred_drop, cPres_test))
  prestest <- transform(prestest,
                        catchment = as.factor(prestest$catchment),
                        lith = as.factor(prestest$lith),
                        marine = as.factor(prestest$marine))
  cBackg_test <- c_bg_all[kfold_back == 1, ]
  abstest <- data.frame(raster::extract(pred_drop, cBackg_test))
  abstest <- transform(abstest,
                       catchment = as.factor(abstest$catchment),
                       lith = as.factor(abstest$lith),
                       marine = as.factor(abstest$marine))
  # GLM with selected variables
  c_glm_selmodel <- glm(pab ~ ., family = binomial(link = "logit"),
                        data = envtrain)}
# Warning message:
#   glm.fit: fitted probabilities numerically 0 or 1 occurred

c_glm_sel_e <- evaluate(p=prestest,a=abstest, c_glm_selmodel)
```

```
# class          : ModelEvaluation
# n presences    : 24
# n absences     : 2000
# AUC            : 0.9146042
# cor           : 0.09460181
# max TPR+TNR at : -4.0666

varSel_auc <- slot(c_glm_sel_e, "auc")
# 0.9146042

## Variable importance (varSel) ----
c_vi_varSel_glm <- vi_model(c_glm_selmodel, type = "stat")
write.csv(c_vi_varSel_glm, "./c_vi_varSel_glmexpcomparison.csv")

save.image("./cGLM_exploratory_comparison.RData")

# Variable selection 2 ----
set.seed(42)
# Run glmulti genetic algorithm 10x
c_glmulti_l1_g <- list()

for (i in 1:10){
  c_glmulti_l1_g[[i]] <- glmulti(c_glm_selmodel,
                                 crit = aicc,
                                 level = 1,
                                 method = "g",
                                 family = binomial,
                                 confsetsize = 50,
                                 plotty = F)
}
saveRDS(c_glmulti_l1_g, "./c_glmulti_exp_comparison.RDS")

# Get AIC values from top models of each glmulti
c_glmulti_aicc <- c_glmulti_l1_g[[1]]@objects[[1]]$aic

for (i in 2:10) {
  c_glmulti_aicc <- cbind(c_glmulti_aicc, c_glmulti_l1_g[[i]]@objects[[1]]$aic)
}
save.image("./cGLM_exploratory_comparison.RData")

## Write top models to csv for all glmulti replicates ----
c.output.list <- list()
c.output <- data.frame(matrix(NA, nrow = 50, ncol = 2)) # Create an empty data.frame
colnames(c.output) <- c("variables", "AICc")
for (i in 1:10){
  for (j in 1:50){
```

```
    formula <- as.character(c_glmulti_l1_g[[i]]@formulas[[j]])
    aicc <- c_glmulti_l1_g[[i]]@objects[[j]]$aic
    c.output[j, 1] <- formula[3]
    c.output[j, 2] <- aicc

  }
  c.output.list[[i]] <- c.output
}
curr.models <- rbind(c.output.list[[1]], c.output.list[[2]], c.output.list[[3]],
                     c.output.list[[4]], c.output.list[[5]], c.output.list[[6]],
                     c.output.list[[7]], c.output.list[[8]], c.output.list[[9]],
                     c.output.list[[10]])
curr.models <- curr.models[order(curr.models$AICc),]

write.csv(curr.models, "./c_glmulti_topmodels.csv")

## Find which layers glmulti dropped ----
print(c_glmulti_aicc)
# [1,]       849.2388 849.2388 849.2388 849.2388 849.2388 849.2388 849.2388 851.3133
849.2388 849.2388
min(c_glmulti_aicc)
# 849.2388
# glmulti 1-10

c_glmulti_l1_g[[1]]@formulas[[1]]
# pab ~ 1 + catchment + lith + marine + clay + elev + for.1922 + for.1924 + for.1927 + for.1928
+ for.1929 + for.1931 + for.1981 + pH + silt


# Original: catchment, clay, elev, for.1920, for.1921, for.1922, for.1924, for.1925, for.1926,
for.1927, for.1928, for.1929, for.1930, for.1931, for.1980, for.1981, lith, marine, pH, silt, slope
# Reduced: catchment, clay, elev, for.1922, for.1924, for.1927, for.1928, for.1929, for.1931,
for.1981, lith, marine, pH, silt
# Dropped: for.1920, for.1921, for.1925, for.1926, for.1930, for.1980, slope

# Drop layers 2 ----
pred_drop2 <- dropLayer(pred_drop, c("for.1920", "for.1921", "for.1925", "for.1926",
"for.1930", "for.1980", "slope"))

save.image("./cGLM_exploratory_comparison.RData")

# GLM with reduced variables  ----
{
  set.seed(42)
  kfold_pres <- kfold(cPres, folds)
  kfold_back <- kfold(c_bg_all, folds)
```

```
  ## Create training data using reduced set of predictors ----
  cPres_train <- cPres[kfold_pres != 1, ]
  cBackg_train <- c_bg_all[kfold_back != 1, ]
  pab_train <- c(rep(1, nrow(cPres_train)), rep(0, nrow(cBackg_train)))
  envtrain <- raster::extract(pred_drop2, rbind(cPres_train,cBackg_train))
  envtrain <- data.frame(cbind(pab=pab_train, envtrain))
  envtrain <- transform(envtrain,
                        catchment = as.factor(envtrain$catchment),
                        lith = as.factor(envtrain$lith),
                        marine = as.factor(envtrain$marine))
  prestrain <- data.frame(raster::extract(pred_drop2, cPres_train))
  prestrain <- transform(prestrain,
                         catchment = as.factor(prestrain$catchment),
                         lith = as.factor(prestrain$lith),
                         marine = as.factor(prestrain$marine))
  abstrain <- data.frame(raster::extract(pred_drop2, cBackg_train))
  abstrain <- transform(abstrain,
                        catchment = as.factor(abstrain$catchment),
                        lith = as.factor(abstrain$lith),
                        marine = as.factor(abstrain$marine))

  ## Create testing data ----
  cPres_test <- cPres[kfold_pres == 1, ]
  prestest <- data.frame(raster::extract(pred_drop2, cPres_test))
  prestest <- transform(prestest,
                        catchment = as.factor(prestest$catchment),
                        lith = as.factor(prestest$lith),
                        marine = as.factor(prestest$marine))
  cBackg_test <- c_bg_all[kfold_back == 1, ]
  abstest <- data.frame(raster::extract(pred_drop2, cBackg_test))
  abstest <- transform(abstest,
                       catchment = as.factor(abstest$catchment),
                       lith = as.factor(abstest$lith),
                       marine = as.factor(abstest$marine))
  ## GLM with reduced variables ----
  c_reduceVar_glm <- glm(pab ~ ., family = binomial(link = "logit"),
                         data = envtrain)

  c_summ <- summary(c_reduceVar_glm)
}


c_summ
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
```

```
# (Intercept) -7.950e+00  6.533e-01 -12.169  < 2e-16 ***
# catchment2   1.142e+00  3.112e-01   3.668 0.000245 ***
# catchment3   2.073e+00  3.746e-01   5.533 3.15e-08 ***
# catchment4   1.940e+00  4.608e-01   4.210 2.56e-05 ***
# catchment5  -1.224e+01  4.399e+02  -0.028 0.977800
# catchment6  -1.332e+01  1.214e+03  -0.011 0.991244
# clay        -5.480e-01  1.883e-01  -2.910 0.003613 **
# elev        -3.734e+00  9.341e-01  -3.997 6.41e-05 ***
# for.1922     2.452e-01  1.197e-01   2.050 0.040393 *
# for.1924    -9.961e-01  3.427e-01  -2.907 0.003651 **
# for.1927     9.833e-01  1.626e-01   6.049 1.46e-09 ***
# for.1928    -1.429e+00  1.002e+00  -1.427 0.153650
# for.1929     1.048e+00  4.247e-01   2.469 0.013556 *
# for.1931     2.289e-01  9.083e-02   2.520 0.011740 *
# for.1981    -3.376e-01  1.711e-01  -1.973 0.048495 *
# lith13       5.924e-03  2.886e-01   0.021 0.983622
# lith14      -1.743e+00  7.129e-01  -2.445 0.014492 *
# lith19      -1.650e+01  1.520e+03  -0.011 0.991340
# lith999     -1.702e+00  4.665e-01  -3.649 0.000264 ***
# marine1     -9.966e-01  3.952e-01  -2.522 0.011673 *
# pH          -3.769e-01  1.305e-01  -2.889 0.003867 **
# silt         3.587e-01  1.652e-01   2.171 0.029968 *
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 1060.03  on 8097  degrees of freedom
# Residual deviance:  805.24  on 8076  degrees of freedom
# AIC: 849.24
#
# Number of Fisher Scoring iterations: 18

write.csv(c_summ$coefficients, "./c_final_glm_summaryexpcomparison.csv")


(c_reduceVar_train_e <- evaluate(p=prestrain, a=abstrain, c_reduceVar_glm))
# class          : ModelEvaluation
# n presences    : 98
# n absences     : 8000
# AUC            : 0.8971964
# cor            : 0.1020383
# max TPR+TNR at : -4.176151

(c_reduceVar_test_e <- evaluate(p=prestest,a=abstest, c_reduceVar_glm))
# class          : ModelEvaluation
```

```
# n presences    : 24
# n absences     : 2000
# AUC            : 0.9110208
# cor            : 0.1001951
# max TPR+TNR at : -4.050003
(cGLM_eval <- ensemble.evaluate(eval =  c_reduceVar_test_e, eval.train =
c_reduceVar_train_e))
# Calculated fixed threshold of -4.17615081 corresponding to highest sum of sensitivity and
specificity
#       AUC       TSS      SEDI  TSS.fixed SEDI.fixed  FNR.fixed  MCR.fixed    AUCdiff
# 0.91102083 0.74166667 0.94919699 0.72566667 0.86531032 0.08333333 0.18972332 -
0.01382440

save.image("./cGLM_exploratory_comparison.RData")

## calculate the AUC for comparison with other models ----

(caucGLM <- slot(c_reduceVar_test_e, "auc"))
# [1] 0.8902292

## GLM thresholds ----
(c_Opt_glm <-  c_reduceVar_test_e@t[which.max(c_reduceVar_test_e@TPR +
c_reduceVar_test_e@TNR)])
# -4.257425

(cthGLM <- plogis(c_Opt_glm)) # threshold on logit scale
# [1] 0.01396105

## Variable importance ----
c_vi_reduceVar_glm <- vi_model(c_reduceVar_glm, type = "stat")

write.csv(c_vi_reduceVar_glm, "./c_vi_reduceVar_glmexpcomparison.csv")

save.image("./cGLM_exploratory_comparison.RData")


  ------------------------------------------------------------------------------------------------------------
----------------------

# 4. Final RF and Maxent of current presence-absence-background data
##############################################################################
#
# Final current models ########################################################
##############################################################################
#
# Using only variables from top model (current pop. Maxent model)
```

```
rm(list = ls()) # remove all objects from R
load("./cFinalRFMx.RData")

# Load libraries ----
library(dismo)
library(dplyr)
library(rasterVis)
library(rgdal)
library(rJava)
library(SDMtune)
library(sp)
library(terra)
library(zeallot)

# Load scaled rasters into terra ----
files <- list.files(path = "./Predictors/sw_ext/Scaled/Final/",
                    pattern = "tif",
                    all.files = TRUE,
                    full.names = TRUE,
                    include.dirs = FALSE)

# Stack rasters in terra ----
predictors <- terra::rast(files)
names(predictors) # check order of layers

names(predictors) <- c("catchment", "clay", "elev", "for.1924", "for.1927",
                       "for.1930", "for.1980",
                       "lith", "pH", "silt", "slope")
# for.1920 = Laurentian-Acadian Northern Hardwoods Forest
# for.1921 = Northeastern Interior Dry-Mesic Oak Forest
# for.1922 = Northern Atlantic Coastal Plain Hardwood Forest
# for.1924 = Laurentian-Acadian Northern Pine(-Oak) Forest
# for.1925 = Laurentian-Acadian Pine-Hemlock-Hardwood Forest
# for.1926 = Central Appalachian Dry Oak-Pine Forest
# for.1927 = Appalachian (Hemlock)-Northern Hardwood Forest
# for.1928 = Acadian Low-Elevation Spruce-Fir-Hardwood Forest
# for.1929 = Acadian-Appalachian Montane Spruce-Fir Forest
# for.1930 = Central Appalachian Pine-Oak Rocky Woodland
# for.1931 = Northern Atlantic Coastal Plain Maritime Forest
# for.1980 = Boreal Jack Pine-Black Spruce Forest
# for.1981 = Northeastern Interior Pine Barrens

# Change categorical variables to factors ----
predictors$catchment <- as.factor(predictors$catchment)
predictors$lith <- as.factor(predictors$lith)
```

```
# predictors$marine <- as.factor(predictors$marine)

# Check rasters
# x11()
# par(mfrow=c(4,4))
# plot(predictors[[1:16]])

# Read in presence/absence points ----
bds <- readOGR("./BDSsites_snapped.shp") # snapped to catchment position raster
bds$cOccu[bds$cOccu == -9999] <- NA # Arc won't let NA's be assigned to numeric columns, so fix this here
bds$hOccu[bds$hOccu == -9999] <- NA
bds$Year_Est[bds$Year_Est == -9999] <- NA
bds$YearSamp[bds$YearSamp == -9999] <- NA

# Confirm that all instances of -9999 have been removed
summary(bds$cOccu) # 7 NA's
summary(bds$hOccu) # 93 NA's, only one confirmed historic absence (other sites unknown)
summary(bds$Year_Est)
#  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
#  1898    2005    2014    2001    2021    2022       6
summary(bds$YearSamp) # 0 refers to unknown sampling year, NA refers to sites not sampled after 2000
#   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
#      0    2014    2021    2008    2022    2022       8
# write.csv(bds, "./bdssites_snapped.csv")

# Current presence coordinates ----
cPresent <- subset(bds, cOccu == 1) # 122 presences
cPres <- cPresent@coords[,1:2]

# Current absence coordinates ----
cAbsent <- subset(bds, cOccu == 0) # 121 absences
cAbs <- cAbsent@coords[,1:2]
c.ab <- 10000-length(cAbsent)

# Generate background coordinates ----
set.seed(42)
# bg_coords <- terra::spatSample(predictors,
#                                size = 50000, # returns 12107 points (many NA's from points on land)
#                                method = "random",
#                                na.rm = TRUE,
#                                xy = TRUE,
#                                values = FALSE)
# saveRDS(bg_coords, "./all_bg_coords_terra.rds") # save for later analyses
bg_coords <- readRDS("./all_bg_coords_terra.rds")
```

```
bg_coords_cab <- bg_coords[1:c.ab,] # keep first 10,000 points for Maxent

# Combine background and absence coordinates
c_bg_all <- rbind(bg_coords_cab, cAbs)
# saveRDS(c_bg_all, "./curr_bgab_terra_all.rds") # save points
save.image(file = "cFinalRFMx.RData")

# Create SWD object ----
c_data <- prepareSWD(species = "BDS",
                     p = cPres, # 122 presence points
                     a = c_bg_all,
                     env = predictors,
                     categorical = c("catchment", "lith"))

# Split data for cross-validation ----
c(ctrain, cval, ctest) %<-% trainValTest(c_data,
                                         val = 0.2,
                                         test = 0.2,
                                         only_presence = FALSE,
                                         seed = 42) # The only_presence argument is used to split only the
presence and not the background locations (Maxent only)

save.image(file = "cFinalRFMx.RData")

# Random forest model----
## Train a model with default settings ----

set.seed(42)
c_default_rf <- train(method = "RF", # can only do classification rf in SDMtune
                      data = ctrain)
# Presence locations: 125
# Absence locations: 8000
# mtry: 3
# ntree: 500
# nodesize: 1
# Continuous: clay elev for.1924 for.1927 for.1930 for.1980 pH silt slope
# Categorical: catchment lith

cat("Training auc: ", auc(c_default_rf))
# Training auc:  1 # overfitting
cat("Training TSS: ", tss(c_default_rf))
# Training TSS:  1
cat("Testing auc: ", auc(c_default_rf, test = cval))
# Testing auc:  0.8849633
cat("Testing TSS: ", tss(c_default_rf, test = cval))
# Testing TSS:  0.6256462
```

```
c_default_rf@model@model$confusion
#      0 1 class.error
# 0 5991 7 0.001167056
# 1   66 7 0.904109589

### Variable importance default ----
(c_vi_defaultrf <- varImp(c_default_rf,
                          permut = 10))
#    Variable Permutation_importance    sd
# 1       silt                 25.8 0.000
# 2   for.1980                  22.8 0.000
# 3   for.1927                  16.4 0.001
# 4       elev                 15.3 0.000
# 5  catchment                   8.4 0.000
# 6      slope                  4.3 0.000
# 7       clay                  3.3 0.000
# 8         pH                  1.9 0.000
# 9   for.1924                  1.6 0.000
# 10  for.1930                  0.2 0.000
# 11      lith                 0.0 0.000
write.csv(c_vi_defaultrf, "./c_vi_default_rffinal.csv")

save.image(file = "cFinalRFMx.RData")

## K-fold cross-validation ----
c_folds_rf <- randomFolds(ctrain,
                          k = 10,
                          only_presence = FALSE,
                          seed = 42)

c_kfold_rf <- train("RF",
                    data = ctrain,
                    folds = c_folds_rf)
# Presence locations: 73
# Absence locations: 5998
# Replicates: 10
# mtry: 3
# ntree: 500
# nodesize: 1

cat("Training AUC: ", auc(c_kfold_rf))
# Training AUC:  1 # overfitting
cat("Testing AUC: ", auc(c_kfold_rf, test = cval))
# Testing AUC:  0.8857971
cat("Testing TSS: ", tss(c_kfold_rf, test = cval))
```

```
# Testing TSS:  0.6479302

### Variable importance kfold ----
(c_vi_kfoldrf <- varImp(c_kfold_rf,
                        permut = 10))
#     Variable Permutation_importance    sd
# 1       silt                24.56 6.157
# 2   for.1927                 21.35 8.206 Appalachian (Hemlock)-Northern Hardwood Forest
# 3   for.1980                 18.61 7.399 Boreal Jack Pine-Black Spruce Forest
# 4       elev                13.80 4.529
# 5  catchment                  9.71 4.079
# 6      slope                 3.96 2.038
# 7       clay                 3.86 1.373
# 8         pH                 2.66 1.067
# 9   for.1924                  1.23 0.618 Laurentian-Acadian Northern Pine(-Oak) Forest
# 10  for.1930                  0.17 0.149 Central Appalachian Pine-Oak Rocky Woodland
# 11      lith                0.05 0.071

write.csv(c_vi_kfoldrf, "./c_vi_kfold_rffinal.csv")

save.image(file = "cFinalRFMx.RData")

## Tune hyperparameters 2-----
getTunableArgs(c_kfold_rf)
# "mtry"     "ntree"    "nodesize"

c_rf <- list(mtry = 1:11,
             ntree = seq(500,2000,200),
             nodesize = 1:15)

c_om_rf <- optimizeModel(c_kfold_rf,
                         hypers = c_rf,
                         metric = "auc",
                         seed = 42)

saveRDS(c_om_rf, "./c_om_rffinal.rds")

c_best_model_rf <- c_om_rf@models[[1]]

c_om_rf@results[1, ]
#   mtry ntree nodesize train_AUC    test_AUC    diff_AUC
# 1    2  1300       10 0.9999966    0.939671  0.06032566

write.csv(c_om_rf@results, "./c_om_rf_results_final.csv")

## Merge SWD ----
```

```
# Index of the best model
c_index_rf <- which.max(c_om_rf@results$test_AUC)

# New train dataset containing only the selected variables
# c_new_train_rf <- c_reduced_var_rf@data

# Merge data
c_merged_data_rf <- mergeSWD(ctrain,
                        cval,
                        only_presence = FALSE)
# Presence locations: 125
# Absence locations: 8000
save.image(file = "cFinalRFMx.RData")

## Final model ----
c_folds_rf2 <- randomFolds(c_merged_data_rf,
                      k = 10,
                      only_presence = FALSE,
                      seed = 42)

c_kfold_rf2 <- train("RF",
                data = c_merged_data_rf,
                folds = c_folds_rf2,
                mtry = c_om_rf@results[c_index_rf, 1], # mtry = 2
                ntree = c_om_rf@results[c_index_rf, 2], # ntree = 1300
                nodesize = c_om_rf@results[c_index_rf, 3]) # nodesize = 10
# Presence locations: 125
# Absence locations: 8000
# Replicates: 10
# mtry: 2
# ntree: 1300
# nodesize: 10
# Continuous: clay elev for.1924 for.1927 for.1930 for.1980 pH silt slope
# Categorical: catchment lith

cat("Training AUC: ", auc(c_kfold_rf2))
# Training AUC:  0.9999966 # overfitting
cat("Testing AUC: ", auc(c_kfold_rf2, test = ctest))
# Testing AUC:  0.9081711
cat("Testing TSS: ", tss(c_kfold_rf2, test = ctest))
# Testing TSS:  0.7283722
caucrf <- auc(c_kfold_rf2, test = ctest) # for weighted mean of models

### Variable importance final ----
(c_vi_finalrf <- varImp(c_kfold_rf2,
                    permut = 10))
```

```
#     Variable Permutation_importance    sd
# 1       silt                26.86 2.553
# 2       elev                18.11 2.837
# 3  catchment                   12.98 2.574
# 4   for.1927                 12.31 2.980
# 5      slope                  8.80 2.477
# 6       clay                 8.38 1.348
# 7         pH                 5.50 0.841
# 8   for.1980                   3.96 1.655
# 9   for.1924                   1.92 0.483
# 10      lith                 0.67 0.254
# 11  for.1930                    0.47 0.206
write.csv(c_vi_finalrf, "./c_vi_final_rf.csv")

## Create distribution map ----
c_map_rf <- predict(c_kfold_rf2,
                    data = predictors,
                    file = "./Predictions/Prob/Current/c_prob_rffinal.tif",
                    overwrite = TRUE)
c_map_rf <- terra::rast("./Predictions/Prob/Current/c_prob_rffinal_mean.tif")
### Thresholds ----
cthsrf <- list()
c_kfold_ths <- matrix(NA, nrow = 10, ncol = 1)
for (i in 1:10){
  cthsrf[[i]] <- thresholds(c_kfold_rf@models[[i]], test = ctest)
  c_kfold_ths[i,1] <- cthsrf[[i]][5,2]
}
saveRDS(cthsrf, "./c_kfold_thresholds_rffinal.rds")

(c_ths_rf <- mean(c_kfold_ths))
# 0.0094

## Presence/absence map ----
c_pamap_rf <- plotPA(c_map_rf,
                     th = c_ths_rf,
                     filename = "./Predictions/PresAbs/Current/c_pamap_rffinal.tif",
                     overwrite = TRUE)
x11()
plot(c_pamap_rf)

# Maxent model----
## Train a model with default settings ----
set.seed(42)
c_default_mx <- train(method = "Maxent",
                      data = ctrain)
```

```
cat("Training auc: ", auc(c_default_mx))
# Training auc:  0.9466831
cat("Training TSS: ", tss(c_default_mx))
# Training TSS:  0.7996296
cat("Testing auc: ", auc(c_default_mx, test = cval))
# Testing auc: 0.885276
cat("Testing TSS: ", tss(c_default_mx, test = cval))
# Testing TSS:  0.6337335

### Variable importance default ----
(c_vi_defaultmx <- varImp(c_default_mx,
                          permut = 10))
#     Variable Permutation_importance    sd
# 1   for.1927                 44.0 0.029
# 2  catchment                  11.6 0.007
# 3   for.1980                 10.4 0.003
# 4       elev                 8.2 0.010
# 5   for.1930                  7.2 0.008
# 6      slope                 5.3 0.007
# 7       lith                3.9 0.004
# 8         pH                 3.9 0.002
# 9       silt                2.4 0.003
# 10      clay                 1.5 0.003
# 11  for.1924                  1.5 0.003

write.csv(c_vi_defaultmx, "./c_vi_default_mxfinal.csv")
save.image(file = "cFinalRFMx.RData")

## K-fold cross-validation ----
c_folds_mx <- randomFolds(ctrain,
                          k = 10,
                          only_presence = FALSE,
                          seed = 42)

c_kfold_mx <- train("Maxent",
                    data = ctrain,
                    folds = c_folds_mx)

cat("Training AUC: ", auc(c_kfold_mx))
# Training AUC:  0.9474427
cat("Testing AUC: ", auc(c_kfold_mx, test = cval))
# Testing AUC:  0.8791146
cat("Testing TSS: ", tss(c_kfold_mx, test = cval))
# Testing TSS:  0.6112577

### Variable importance kfold ----
```

```
(c_vi_kfoldmx <- varImp(c_kfold_mx,
                        permut = 10))
#     Variable Permutation_importance    sd
# 1   for.1927                 43.18 2.709
# 2  catchment                  11.25 1.715
# 3   for.1980                  9.91 1.088
# 4       elev                 8.98 1.781
# 5   for.1930                  7.50 1.715
# 6      slope                 5.75 1.831
# 7         pH                 3.70 1.080
# 8       lith                3.57 0.691
# 9       silt                2.70 0.782
# 10      clay                  2.07 1.254
# 11  for.1924                  1.37 0.424

write.csv(c_vi_kfoldmx, "./c_vi_kfold_mxfinal.csv")

save.image(file = "cFinalRFMx.RData")

## Tune hyperparameters-----
getTunableArgs(c_kfold_mx)
# [1] "fc"   "reg"  "iter"

c_mx <- list(fc = c("l", "lq", "lh", "lqp", "lqph", "lqpht"),
             iter = seq(300,1100,200),
             reg = seq(0.2,1,0.1))

c_om_mx <- optimizeModel(c_kfold_mx,
                         hypers = c_mx,
                         metric = "auc",
                         seed = 42)

c_best_model_mx <- c_om_mx@models[[1]]

c_om_mx@results[1, ]
#   fc    reg iter train_AUC  test_AUC   diff_AUC
# 1 lqpht 0.3  300 0.9737843 0.9276561 0.04612822

save.image(file = "cFinalRFMx.RData")
write.csv(c_om_mx@results, "./c_om_mx_resultsfinal.csv")

## Merge SWD ----
# Index of the best model
c_index_mx <- which.max(c_om_mx@results$test_AUC)

# New train dataset containing only the selected variables
```

```
# c_new_train_mx <- c_reduced_var_mx@data

# Merge only presence data
c_merged_data_mx <- mergeSWD(ctrain,
                             cval,
                             only_presence = FALSE)
# Presence locations: 125
# Absence locations: 8000

## Final model ----
c_folds_mx2 <- randomFolds(c_merged_data_mx,
                           k = 10,
                           only_presence = FALSE,
                           seed = 42)

c_kfold_mx2 <- train("Maxent",
                     data = c_merged_data_mx,
                     folds = c_folds_mx2,
                     fc = c_om_mx@results[c_index_mx, 1], # fc = lh
                     reg = c_om_mx@results[c_index_mx, 2], # reg = 0.5
                     iter = c_om_mx@results[c_index_mx, 3]) # iter = 300
# Presence locations: 97
# Absence locations: 7997
# Replicates: 10
# lh: lqpht
# reg: 0.3
# iter: 300

cat("Training AUC: ", auc(c_kfold_mx2))
# Training AUC:  0.9638511 # overfitting
cat("Testing AUC: ", auc(c_kfold_mx2, test = ctest))
# Testing AUC:  0.9212166
cat("Testing TSS: ", tss(c_kfold_mx2, test = ctest))
# Testing TSS:  0.7451446
caucmx <- auc(c_kfold_mx2, test = ctest) # for weighted mean of models

### Variable importance final ----
(c_vi_finalmx <- varImp(c_kfold_mx2,
                        permut = 10))
#     Variable Permutation_importance    sd
# 1   for.1927                 54.03 4.373
# 2       elev                11.15 1.883
# 3      slope                 8.01 1.084
# 4  catchment                    6.02 1.146
# 5   for.1930                  4.49 1.038
# 6       lith                3.72 1.536
```

```
# 7   for.1924                3.57 1.277
# 8       silt                3.09 0.638
# 9       clay                2.60 0.999
# 10  for.1980                 2.54 0.789
# 11        pH                0.77 0.313
write.csv(c_vi_finalmx, "./c_vi_final_mx.csv")
write.csv(c_kfold_mx2@model@lambdas, "./c_final_mx_lambdas.csv")

## Create distribution map ----
c_map_mx <- predict(c_kfold_mx2,
                    data = predictors,
                    file = "./Predictions/Prob/Current/c_prob_mxfinal.tif",
                    overwrite = TRUE)
c_map_mx <- terra::rast("./Predictions/Prob/Current/c_prob_mxfinal_mean.tif")
### Thresholds ----
cthsmx <- list()
c_kfold_ths_mx <- matrix(NA, nrow = 10, ncol = 1)
for (i in 1:10){
  cthsmx[[i]] <- thresholds(c_kfold_mx2@models[[i]], test = ctest)
  c_kfold_ths_mx[i,1] <- cthsmx[[i]][5,2]
}
saveRDS(cthsmx, "./c_thresholds_mxfinal.rds")

(c_ths_mx <- mean(c_kfold_ths_mx))
# 0.1414215

## Presence/absence map ----
c_pamap_mx <- plotPA(c_map_mx,
                     th = c_ths_mx,
                     filename = "./Predictions/PresAbs/Current/c_pamap_mxfinal.tif",
                     overwrite = TRUE)
x11()
plot(c_pamap_mx)
save.image(file = "cFinalRFMx.RData")


-------------------------------------------------------------------------------------------------------------
--------------------

# 5. Final GLM of current presence-absence-background data
#############################################################################
# Current presence-absence-background data#####
#############################################################################
#load("cGLMfinal.RData")

rm(list = ls())
```

```
# Load libraries ----
library(BiodiversityR)
library(caret)
library(dismo)
library(dplyr)
library(glmulti)
library(raster)
library(rgdal)
library(tidyverse)
library(vip)

# Load scaled rasters into terra ----
files <- list.files(path = "./Predictors/sw_ext/Scaled/Final/",
                    pattern = "tif",
                    all.files = TRUE,
                    full.names = TRUE,
                    include.dirs = FALSE)

# Stack rasters in terra ----
predictors <- terra::rast(files)
names(predictors) # check order of layers

names(predictors) <- c("catchment", "clay", "elev", "for.1924", "for.1927",
                       "for.1930", "for.1980",
                       "lith", "pH", "silt", "slope")
# for.1920 = Laurentian-Acadian Northern Hardwoods Forest
# for.1921 = Northeastern Interior Dry-Mesic Oak Forest
# for.1922 = Northern Atlantic Coastal Plain Hardwood Forest
# for.1924 = Laurentian-Acadian Northern Pine(-Oak) Forest
# for.1925 = Laurentian-Acadian Pine-Hemlock-Hardwood Forest
# for.1926 = Central Appalachian Dry Oak-Pine Forest
# for.1927 = Appalachian (Hemlock)-Northern Hardwood Forest
# for.1928 = Acadian Low-Elevation Spruce-Fir-Hardwood Forest
# for.1929 = Acadian-Appalachian Montane Spruce-Fir Forest
# for.1930 = Central Appalachian Pine-Oak Rocky Woodland
# for.1931 = Northern Atlantic Coastal Plain Maritime Forest
# for.1980 = Boreal Jack Pine-Black Spruce Forest
# for.1981 = Northeastern Interior Pine Barrens

# Change categorical variables to factors ----
```

```
predictors$catchment <- as.factor(predictors$catchment)
predictors$lith <- as.factor(predictors$lith)
# predictors$marine <- as.factor(predictors$marine)

# Check rasters
# x11()
# par(mfrow=c(4,4))
# plot(predictors[[1:16]])

# Read in presence/absence points ----
bds <- readOGR("./BDSsites_snapped.shp") # snapped to catchment position raster
bds$cOccu[bds$cOccu == -9999] <- NA # Arc won't let NA's be assigned to numeric columns, so fix this here
bds$hOccu[bds$hOccu == -9999] <- NA
bds$Year_Est[bds$Year_Est == -9999] <- NA
bds$YearSamp[bds$YearSamp == -9999] <- NA

# Confirm that all instances of -9999 have been removed
summary(bds$cOccu) # 7 NA's
summary(bds$hOccu) # 93 NA's, only one confirmed historic absence (other sites unknown)
summary(bds$Year_Est)
#  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
#  1898    2005    2014    2001    2021    2022       6
summary(bds$YearSamp) # 0 refers to unknown sampling year, NA refers to sites not sampled after 2000
#   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
#      0    2014    2021    2008    2022    2022       8
# write.csv(bds, "./bdssites_snapped.csv")

# Current presence coordinates ----
cPresent <- subset(bds, cOccu == 1) # 121 known presences
cPres <- cPresent@coords[,1:2]

# Current absence coordinates ----
cAbsent <- subset(bds, cOccu == 0) # 120 known absences
cAbsent <- subset(cAbsent, Site_Name != "MARRPD-01") # Remove absences outside of sw extent
cAbsent <- subset(cAbsent, Site_Name != "MARRPD-02")
cAbsent <- subset(cAbsent, Site_Name != "MARBRK")
cAbsent <- subset(cAbsent, Site_Name != "BROBRK")
```

```
cAbs <- cAbsent@coords[,1:2]
c.ab <- 10000-length(cAbsent)

# Read in background coordinates ----
bg_coords <- readRDS("./all_bg_coords_terra.rds") # use same background points as with
MaxEnt and RF
bg_coords_cab <- bg_coords[1:c.ab,] # keep first 10,000 points

# Combine background and absence coordinates ----
c_bg_all <- rbind(bg_coords_cab, cAbs)

# Default GLM ----
## Partition presence and absence data ----
folds <- 5 # percentage split (80/20) as in other models

set.seed(42)
kfold_pres <- kfold(cPres, folds)
kfold_back <- kfold(c_bg_all, folds)

# Create training data
cPres_train <- cPres[kfold_pres != 1, ]
cBackg_train <- c_bg_all[kfold_back != 1, ]
pab_train <- c(rep(1, nrow(cPres_train)), rep(0, nrow(cBackg_train)))
envtrain <- raster::extract(predictors, rbind(cPres_train,cBackg_train))
envtrain <- data.frame(cbind(pab=pab_train, envtrain))
envtrain <- transform(envtrain,
                      catchment = as.factor(envtrain$catchment),
                      lith = as.factor(envtrain$lith))
# Create testing data
cPres_test <- cPres[kfold_pres == 1, ]
prestest <- data.frame(raster::extract(predictors, cPres_test))
prestest <- transform(prestest,
                      catchment = as.factor(prestest$catchment),
                      lith = as.factor(prestest$lith))
cBackg_test <- c_bg_all[kfold_back == 1, ]
abstest <- data.frame(raster::extract(predictors, cBackg_test))
abstest <- transform(abstest,
                     catchment = as.factor(abstest$catchment),
                     lith = as.factor(abstest$lith))
# GLM with all variables
```

```
c_default_glm <- glm(pab ~ ., family = binomial(link = "logit"),
                     data = envtrain)
# Warning message:
#   glm.fit: fitted probabilities numerically 0 or 1 occurred

# Test AUC
c_default_e <- evaluate(p=prestest,a=abstest, c_default_glm)
(default_auc <- slot(c_default_e, "auc"))
# 0.894

## Variable importance ----
c_vi_default_glm <- vi_model(c_default_glm, type = "stat") # Z-statistic
write.csv(c_vi_default_glm, "./c_vi_default_glmfinal.csv")

save.image("./cGLMfinal.RData")

# K-fold cross-validation  ----
# Set random seeds
set.seed(42)
(seeds <- sample.int(1000, 10))
# [1] 561 997 321 153  74 228 146 634  49 128

# Set number of folds
folds <- 5 # percentage split (80% training /20% testing) as in other models

# Create empty outputs to hold results
c_kfold_glm <- list()
c_kfold_e_train <- list()
c_kfold_e_test <- list()
kfold_auc_train <- matrix(NA, nrow = 1, ncol = 1)
kfold_auc_test <- matrix(NA, nrow = 1, ncol = 1)
coutput <- data.frame(matrix(NA, nrow = 10, ncol = 3)) # Create an empty data.frame
colnames(coutput) <- c("seed", "trainAUC", "testAUC")

# Cross-validation
for (i in seq_along(seeds)){
  set.seed(i) # set random seed
  kfold_pres <- kfold(cPres, folds) # partition presence and background data according to the
random seed
  kfold_back <- kfold(c_bg_all, folds)
```

```
# Create training data with 4/5 folds
cPres_train <- cPres[kfold_pres != 1, ]
cBackg_train <- c_bg_all[kfold_back != 1, ]
pab_train <- c(rep(1, nrow(cPres_train)), rep(0, nrow(cBackg_train)))
envtrain <- raster::extract(predictors, rbind(cPres_train,cBackg_train))
envtrain <- data.frame(cbind(pab=pab_train, envtrain))
envtrain <- transform(envtrain,
                      catchment = as.factor(envtrain$catchment),
                      lith = as.factor(envtrain$lith))

# Create training data with 4/5 folds
prestrain <- data.frame(raster::extract(predictors, cPres_train))
prestrain <- transform(prestrain,
                       catchment = as.factor(prestrain$catchment),
                       lith = as.factor(prestrain$lith))
abstrain <- data.frame(raster::extract(predictors, cBackg_train))
abstrain <- transform(abstrain,
                      catchment = as.factor(abstrain$catchment),
                      lith = as.factor(abstrain$lith))

# Create testing data with 1/5 folds
cPres_test <- cPres[kfold_pres == 1, ]
prestest <- data.frame(raster::extract(predictors, cPres_test))
prestest <- transform(prestest,
                      catchment = as.factor(prestest$catchment),
                      lith = as.factor(prestest$lith))
cBackg_test <- c_bg_all[kfold_back == 1, ]
abstest <- data.frame(raster::extract(predictors, cBackg_test))
abstest <- transform(abstest,
                     catchment = as.factor(abstest$catchment),
                     lith = as.factor(abstest$lith))

# GLM with all variables
c_kfold_glm[[i]] <- glm(pab ~ ., family = binomial(link = "logit"),
                        data = envtrain)
# Train AUC
c_kfold_e_train[[i]] <- evaluate(p=prestrain, a=abstrain, c_kfold_glm[[i]])
kfold_auc_train[i] <- slot(c_kfold_e_train[[i]], "auc")
```

```
  # Test AUC
  c_kfold_e_test[[i]] <- evaluate(p=prestest,a=abstest, c_kfold_glm[[i]])
  kfold_auc_test[i] <- slot(c_kfold_e_test[[i]], "auc")

  # Populate output data frame
  coutput[i, 1] <- seeds[i]
  coutput[i, 2] <- kfold_auc_train[i]
  coutput[i, 3] <- kfold_auc_test[i]
}
# Warning messages:
#   1: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 2: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 3: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 4: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 5: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 6: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 7: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 8: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 9: glm.fit: fitted probabilities numerically 0 or 1 occurred
# 10: glm.fit: fitted probabilities numerically 0 or 1 occurred
saveRDS(coutput, "./c_kfold_glm_outputfinal.RDS")

coutput
#    seed  trainAUC   testAUC
# 1   561 0.8822666 0.8715000
# 2   997 0.8845880 0.8642708
# 3   321 0.8848329 0.8810625
# 4   153 0.8786454 0.8898542
# 5    74 0.8780281 0.8814167
# 6   228 0.8783673 0.8896042
# 7   146 0.8830969 0.8494583
# 8   634 0.8843839 0.8574583
# 9    49 0.8809094 0.8716458
# 10  128 0.9028202 0.8048542

## calculate the mean AUC ----
(aucGLM_train <- sapply(c_kfold_e_train, function(x){slot(x, 'auc')}))
# [1] 0.8822666 0.8845880 0.8848329 0.8786454 0.8780281 0.8783673 0.8830969 0.8843839
0.8809094 0.9028202
(mean(aucGLM_train))
```

```
# [1] 0.8837939
(aucGLM_test <- sapply(c_kfold_e_test, function(x){slot(x, 'auc')}))
# [1] 0.8715000 0.8642708 0.8810625 0.8898542 0.8814167 0.8896042 0.8494583 0.8574583
0.8716458 0.8048542
(caucGLM <- mean(aucGLM_test))
# 0.8661125

## Thresholds ----
c_kfold_t <- data.frame(matrix(NA, nrow = 10, ncol = 2)) # Create an empty data.frame
colnames(c_kfold_t) <- c("threshold", "logit")

for (i in 1:10) {
  c_kfold_t[i, 1] <- c_kfold_e_test[[i]]@t[which.max(c_kfold_e_test[[i]]@TPR +
c_kfold_e_test[[i]]@TNR)]
  c_kfold_t[i, 2] <- plogis(c_kfold_t[i, 1])
}

c_kfold_t
#    threshold       logit
# 1  -3.520573 0.028732512
# 2  -4.679280 0.009200265
# 3  -4.302436 0.013354787
# 4  -4.255326 0.013989968
# 5  -4.286985 0.013559909
# 6  -4.262792 0.013887348
# 7  -3.911335 0.019621067
# 8  -4.744384 0.008625373
# 9  -3.690804 0.024344484
# 10 -4.478065 0.011227867

mean(c_kfold_t$threshold)
# -4.213198

(cthsglm <- mean(c_kfold_t$logit))
# 0.01565436

## TSS ----
eval <- list()
kfold_tss <- matrix(NA, nrow = 1, ncol = 1)
for (i in 1:10){
```

```
  eval[[i]] <- ensemble.evaluate(eval = c_kfold_e_test[[i]], eval.train = c_kfold_e_train[[i]])
  kfold_tss[i] <- eval[[i]][2]
}
kfold_tss
# [1] 0.6395000 0.5885000 0.6600000 0.7116667 0.6525000 0.6388333 0.5573333 0.5825000
0.6696667 0.5205000
mean(kfold_tss)
# [1] 0.6221

## Variable importance ----
c_vi_kfold_glm <- vi_model(c_kfold_glm[[1]], type = "stat") # Z-statistic

c_vi_kfold_glm <- c_vi_kfold_glm[order(c_vi_kfold_glm$Variable),] %>% select(-Sign)

for (i in 2:length(c_kfold_glm)) {
  vi <- vi_model(c_kfold_glm[[i]], type = "stat")
  c_vi_kfold_glm <- c_vi_kfold_glm %>% left_join(., vi, by = "Variable") %>% select(-Sign)
}
# average ranks together
c_vi_kfold_glm$Avg <- rowMeans(c_vi_kfold_glm[, 2:11])
c_vi_kfold_glm <- c_vi_kfold_glm[order(-c_vi_kfold_glm$Avg),]
write.csv(c_vi_kfold_glm, "./c_vi_kfold_glmfinal.csv")

# average z-scores with sign
ztable <- data.frame(matrix(NA, nrow = 19, ncol = 10))
colnames(ztable) <- c(1:10)
rownames(ztable) <- names(c_kfold_glm[[1]]$coefficients)
for (i in 1:10){
c_summary <- summary(c_kfold_glm[[i]])
coeff <- data.frame(c_summary$coefficients)
ztable[,i] <- coeff$z.value
}
write.csv(ztable, "./c_ztable.csv")



save.image("./cGLMfinal.RData")

# Predict values ----
## Probabilistic ----
```

```
# Rasters on linear scale
for (i in 1:10){
  pglm <- predict(predictors, c_kfold_glm[[i]])
  writeRaster(pglm, paste0(paste0("./Predictions/scratch_glm/pglm_clinear_scale_", i), ".tif"),
overwrite = TRUE)
}
# pglm <- predict(predictors, c_kfold_glm[[9]])
# writeRaster(pglm, "./Predictions/scratch_glm/pglm_clinear_scale_9.tif", overwrite = TRUE)

# Convert rasters to logit scale ----
files <- list.files(path = "./Predictions/scratch_glm/",
                    pattern = "clinear",
                    all.files = TRUE,
                    full.names = TRUE,
                    include.dirs = FALSE)

linrasters <- raster::stack(files)
pglm.mean <- calc(linrasters, fun = mean)
pglm.logit <- calc(pglm.mean, fun = plogis)
writeRaster(pglm.logit, "./Predictions/Prob/Current/c_prob_glmfinal_mean.tif", overwrite =
TRUE)

## Binary presence/absence raster ----
pglm.logit <- raster("./Predictions/Prob/Current/c_prob_glmfinal_mean.tif")
m <- c(0, cthsglm, 0,  cthsglm, 1, 1)
rclmat <- matrix(m, ncol=3, byrow=TRUE)
pglm.pa <- reclassify(pglm.logit, rclmat, include.lowest = TRUE, right = TRUE)
x11()
plot(pglm.pa)
writeRaster(pglm.pa, "./Predictions/PresAbs/Current/c_pamap_glmfinal.tif", overwrite = TRUE)
save.image("./cGLMfinal.RData")


-------------------------------------------------------------------------------------------------------------------
--------------------

# 6. Ensemble models

# Load individual models ----
load("./hFinalRFMx.RData") # historic pops RF and Maxent
load("./cFinalRFMx.RData") # current pops RF and Maxent
load("./cGLMfinal.RData") # current pops GLM
```

```
load("./hGLMfinal.RData") # historic pops GLM


# Load libraries ----
library(dismo)
library(dplyr)
library(rasterVis)
library(rgdal)
library(rJava)
library(SDMtune)
library(sp)
library(terra)
library(zeallot)

# Historic populations ----
## Stack rasters ----
files <- list.files(path = "./Predictions/Prob/Historic/",
                    pattern = "tif",
                    all.files = TRUE,
                    full.names = TRUE,
                    include.dirs = FALSE)
h_prob_predict <- terra::rast(files)
names(h_prob_predict)
names(h_prob_predict) <- c("glm", "mx", "rf")

## Weighted means ----
h_prob_auc <- c(haucGLM, haucmx, haucrf)
# [1] 0.8471952 0.8753726 0.9023403
h_w <- (h_prob_auc-0.5)^2 # subtract 0.5 (the random expectation) and square the result to give
further weight to higher AUC values
# [1] 0.1205445 0.1409046 0.1618777
h_prob_mean <- terra::weighted.mean(h_prob_predict, h_w)

## Mean threshold values ----
h_th_mean <- mean(c(hthsglm, h_ths_mx, h_ths_rf))
# [1] 0.05102805

## Write raster ----
x11()
par(mfrow=c(1,1))
plot(h_prob_mean)

writeRaster(h_prob_mean, "./Predictions/Prob/Historic/h_prob_weighted_mean.tif")
h_pamap_mean <- plotPA(h_prob_mean,
                       th = h_th_mean,
                       filename = "./Predictions/PresAbs/Historic/h_pamap_mean.tif",
```

```
                overwrite = TRUE)
plot(h_pamap_mean)

# Current populations ----
## Stack rasters ----
files <- list.files(path = "./Predictions/Prob/Current/",
                pattern = "tif",
                all.files = TRUE,
                full.names = TRUE,
                include.dirs = FALSE)
c_prob_predict <- terra::rast(files)
names(c_prob_predict)
names(c_prob_predict) <- c("glm", "mx", "rf")

## Weighted mean of model AUCs ----
c_prob_auc <- c(caucGLM, caucmx, caucrf)
# [1] 0.8661125 0.9212166 0.9081711
c_w <- (c_prob_auc-0.5)^2 # subtract 0.5 (the random expectation) and square the result to give
further weight to higher AUC values
# [1] 0.1340384 0.1774234 0.1666036

## Mean threshold values ----
c_th_mean <- mean(c(cthsglm, c_ths_mx, c_ths_rf))
# [1] 0.05549196

## Write rasters ----
c_prob_mean <- terra::weighted.mean(c_prob_predict, c_w)
x11()
par(mfrow=c(1,1))
plot(c_prob_mean)
writeRaster(c_prob_mean, "./Predictions/Prob/Current/c_prob_weighted_mean.tif")
c_pamap_mean <- plotPA(c_prob_mean,
                th = c_th_mean,
                filename = "./Predictions/PresAbs/Current/c_pamap_mean.tif",
                overwrite = TRUE)
plot(c_pamap_mean)

# Change over time ----
c_prob_mean <- rast("./Predictions/Prob/Current/c_prob_weighted_mean.tif")
h_prob_mean <- rast("./Predictions/Prob/Historic/h_prob_weighted_mean.tif")
c_pamap_mean <- rast("./Predictions/PresAbs/Current/c_pamap_mean.tif")
h_pamap_mean <- rast("./Predictions/PresAbs/Historic/h_pamap_mean.tif")


change_in_prob <- c_prob_mean - h_prob_mean
x11()
```

```
plot(change_in_prob)
writeRaster(change_in_prob, "./Predictions/Prob/change_in_prob_presence.tif")

pa_change <- c_pamap_mean - h_pamap_mean
x11()
plot(pa_change)
writeRaster(pa_change, "./Predictions/PresAbs/pa_change.tif")

c_area <- readOGR("./Predictions/PresAbs/Current/c_pamap_area.shp")
h_area <- readOGR("./Predictions/PresAbs/Historic/h_pamap_area.shp")

(loss <- h_area@data$Area - c_area@data$Area)
# 676.033 km2 - 330.807 km2 = 345.226 km2

(prop.left <- c_area@data$Area / h_area@data$Area)
# 0.4893356 range left (continued predicted presence in 49% of historic range)

(prop.lost <- 1 - (c_area@data$Area / h_area@data$Area))
# 0.5106644 range lost (range has shrunk by 51%)

## Maine ----
c_area_ME <- readOGR("./Predictions/PresAbs/Current/c_pamap_area_ME.shp")
h_area_ME <- readOGR("./Predictions/PresAbs/Historic/h_pamap_area_ME.shp")

(loss <- h_area_ME@data$Area - c_area_ME@data$Area)
# 218.585 km2 - 82.346 km2 = 136.239 km2

(prop.left <- c_area_ME@data$Area / h_area_ME@data$Area)
# 0.376723 range left (continued predicted presence in 38% of historic Maine range)

(prop.lost <- 1 - (c_area_ME@data$Area / h_area_ME@data$Area))
# 0.623277 range lost (range has shrunk by 62% in Maine)

## New Hampshire ----
c_area_NH <- readOGR("./Predictions/PresAbs/Current/c_pamap_area_NH.shp")
h_area_NH <- readOGR("./Predictions/PresAbs/Historic/h_pamap_area_NH.shp")

(loss <- h_area_NH@data$Area - c_area_NH@data$Area)
# 457.206 km2 - 248.404 km2 = 208.802 km2

(prop.left <- c_area_NH@data$Area / h_area_NH@data$Area)
# 0.5433087 range left (continued predicted presence in 54% of historic NH range)

(prop.lost <- 1 - (c_area_NH@data$Area / h_area_NH@data$Area))
# 0.4566913 range lost (range has shrunk by 46% in NH)
```

BIOGRAPHY OF THE AUTHOR

Lara Katz was born in Washington, D.C., and grew up in Arlington, Virginia. She was fascinated by wildlife, the outdoors, and conservation from a young age, and spent her summers observing deer and catching hermit crabs near her grandparents' house. Her first camping experience was through a three-week backpacking and canoeing trip with Outward Bound, which introduced her to the Penobscot River and Maine. She graduated from Yorktown High School in 2010, briefly attended Bryn Mawr College, then transferred to the University of Maine in 2011. Lara was first exposed to fisheries science as an undergraduate, where she spent three summers studying stream recolonization by sea lamprey and freshwater stream fish after a dam removal. After graduating with her bachelor's in 2015, Lara spent four years working for Acadia National Park as a wildlife technician and environmental compliance assistant. Lara became especially interested in monitoring rare species as she surveyed Acadia's bats, bumble bees, and wintering seabirds. Her interests in rare species monitoring, geographic information systems, and wetlands led her to return to the University of Maine to pursue her master's degree researching the bridle shiner. After receiving her degree, Lara will begin a Ph.D. program working with a very different and less elusive species, the wild turkey. Lara is a candidate for the Master of Science degree in Wildlife Ecology from the University of Maine in August 2023.