GAP ANALYSIS: A PROPOSED METHODOLOGY TO DESCRIBE AND MAP HISTORICAL AND CONTEMPORARY POPULATIONS AND HABITATS

RESTORING
Great Lakes
COREGONINES

GAP ANALYSIS SCIENCE TEAM

Cory Brant (Co-Lead, U.S. Geological Survey)
Karen Alofs (Co-Lead, University of Michigan)
Chris Castiglione (U.S. Fish and Wildlife Service)
Susan Doka (Fisheries and Oceans Canada)
Alexander Duncan (Chippewas of Nawash Unceded First Nation)
Dave Fielder (Michigan Department of Natural Resources)
Matt Herbert (The Nature Conservancy)
Arunas Liskauskus (Ontario Ministry of Natural Resources and Forestry)
Ed Rutherford (National Oceanic and Atmospheric Administration)
Jason Smith (Sault Ste Marie Tribe of Chippewa Indians)
Ralph Tingley (U.S. Geological Survey)
Ted Treska (U.S. Fish and Wildlife Service)
Ben Turschak (Michigan Department of Natural Resources)
Cindy Chu (Fisheries and Oceans Canada)
Pete Esselman (U.S. Geological Survey)

Citation: Brant, C., C. Castiglione, S. Doka, A. Duncan, D. Fielder, M. Herbert, A. Liskauskus, E. Rutherford, J. Smith, R. Tingley, T. Treska, B. Turschak, C. Chu, P. Esselman, and K. Alofs. 2023. Gap analysis: a proposed methodology to describe and map historical and contemporary populations and habitats. Available from https://www.greatlakesciscoes.org/restoration-framework/planning/

Introduction:

Coregonine populations have generally declined over the past century throughout the Great Lakes, with some species (or forms) now considered extirpated, and some considered extinct (Eshenroder et al. 2016). The Council of Lake Committees (CLC: http://www.glfc.org/council-of-lake-committees.php) endorsed a knowledge and science-based approach to restoring these culturally, ecologically, and economically important fishes in May 2018 (see Bunnell et al. 2023). The approach is often referred to as the Coregonine Restoration Framework, herein abbreviated the CRF (Figure 1). Although the CLC has endorsed this approach, it is important to acknowledge that there is no First Nations representation on the CLC. By including First Nations from the conception of future initiatives like the CRF, the resulting framework would benefit by including both Western science and Indigenous Ecological Knowledge and buy-in from all agencies, Tribes, and First Nations would be enhanced. The CRF is broken into four main phases (Planning, Restore, Evaluate & Learn, and Adjust). Within the Planning phase includes four “Tasks” (orange boxes in Figure 1). Task 1 is to resolve taxonomy, Task 2 is to conduct a Gap Analysis, defined in the CRF as describing and mapping historical and contemporary populations and habitats, Task 3 is to conduct population viability analyses, and Task 4 is to conduct a threats assessment (Figure 1). Science teams assembled for each Task with a charge to review literature and determine the most scientifically defensible methodology for their respective Task. Phase 2 is the implementation of each Task methodology upon review and approval by the Great Lakes Fishery Commission Joint Strategic Plan committees.

Each science team has two co-leads to facilitate meetings, coordinate knowledge transfer, and guide methodology development and writing. Virtual meetings with all team co-leads (herein called Superteam meetings) occur regularly (~monthly) and are facilitated by the Coregonine Science Program Director (Bunnell). Superteam meetings ensure efforts are synergized across teams and not duplicated. In early Superteam meetings, Alofs and Brant, team co-leads for Task 2: Gap Analysis, presented a review on how Gap Analysis methodology might be defined in the context of the CRF and the literature. Regular Superteam discussions led to the Gap Analysis methodology being organized into three subtasks—Subtask I: Collect database sources related to coregonine occurrence, spawning, and habitat (spatial and temporal, historic and contemporary); Subtask II: Model, map, and evaluate species-habitat distributions; and Subtask III: Conduct the “Gap Analysis” to compare historical and contemporary occurrence, spawning areas, and habitats.

The team for Task 2 Gap Analysis (herein, Team 2) formed in late-June 2021 and consists of experts from multiple agencies around the Great Lakes, including:
Cory Brant (Co-Lead, U.S. Geological Survey)
Karen Alofs (Co-Lead, University of Michigan)
Chris Castiglione (U.S. Fish and Wildlife Service)
Susan Doka (Fisheries and Oceans Canada)

Alexander Duncan (Chippewas of Nawash Unceded First Nation)
Dave Fielder (Michigan Department of Natural Resources)
Matt Herbert (The Nature Conservancy)
Arunas Liskauskus (Ontario Ministry of Northern Development, Mines, Natural Resources and Forestry)
Ed Rutherford (National Oceanic and Atmospheric Administration)
Jason Smith (Sault Ste Marie Tribe of Chippewa Indians)
Ralph Tingley (U.S. Geological Survey)
Ted Treska (U.S. Fish and Wildlife Service)
Ben Turschak (Michigan Department of Natural Resources)
Cindy Chu (Fisheries and Oceans Canada)
Pete Esselman (U.S. Geological Survey)

Team 2 discussed and approved the three Gap Analysis subtasks, collected and reviewed literature, held regular virtual meetings, and established co-writing and document sharing arrangements (via Google Drive). Each team member provided a high level of expertise, and, together, through an iterative process, drafted this methodology.

Here, Team 2 presents Phase 1 of Task 2—a Gap Analysis methodology for the CRF. The following methodology is organized into three Subtasks in what we recommend being an iterative process (Figure 2). *Subtask I* covers approaches to gather, digitize, curate, and perform quality assurance and control for historic and contemporary coregonine occurrence, spawning, and habitat data, as well as existing data layers related to habitat for coregonines. This subtask focuses on an approach for developing a database, ensuring accessibility of data and metadata, and ensuring data useability across teams in the CRF. *Subtask II* discusses methods for modeling, mapping, and evaluating species-habitat distributions. In Subtask II we consider how to define historic versus contemporary periods for coregonines in the Great Lakes, review approaches for building species distribution models that can be used to clarify key habitat features and lastly, map suitable habitat. We then present methods for mapping coregonine occurrence and habitat in both periods. In *Subtask III* we discuss methods for comparing historical and contemporary populations and utilized habitats to identify key changes in populations and habitats to help managers make decisions. Each subtask is described in four sections: A. Objectives, B. Considerations, C. Review of Approaches, and D. Recommended Approaches. This methodology was developed at a basin-wide scope yet has the capacity to be consistently applied at the lake level or appropriate ecological scale.

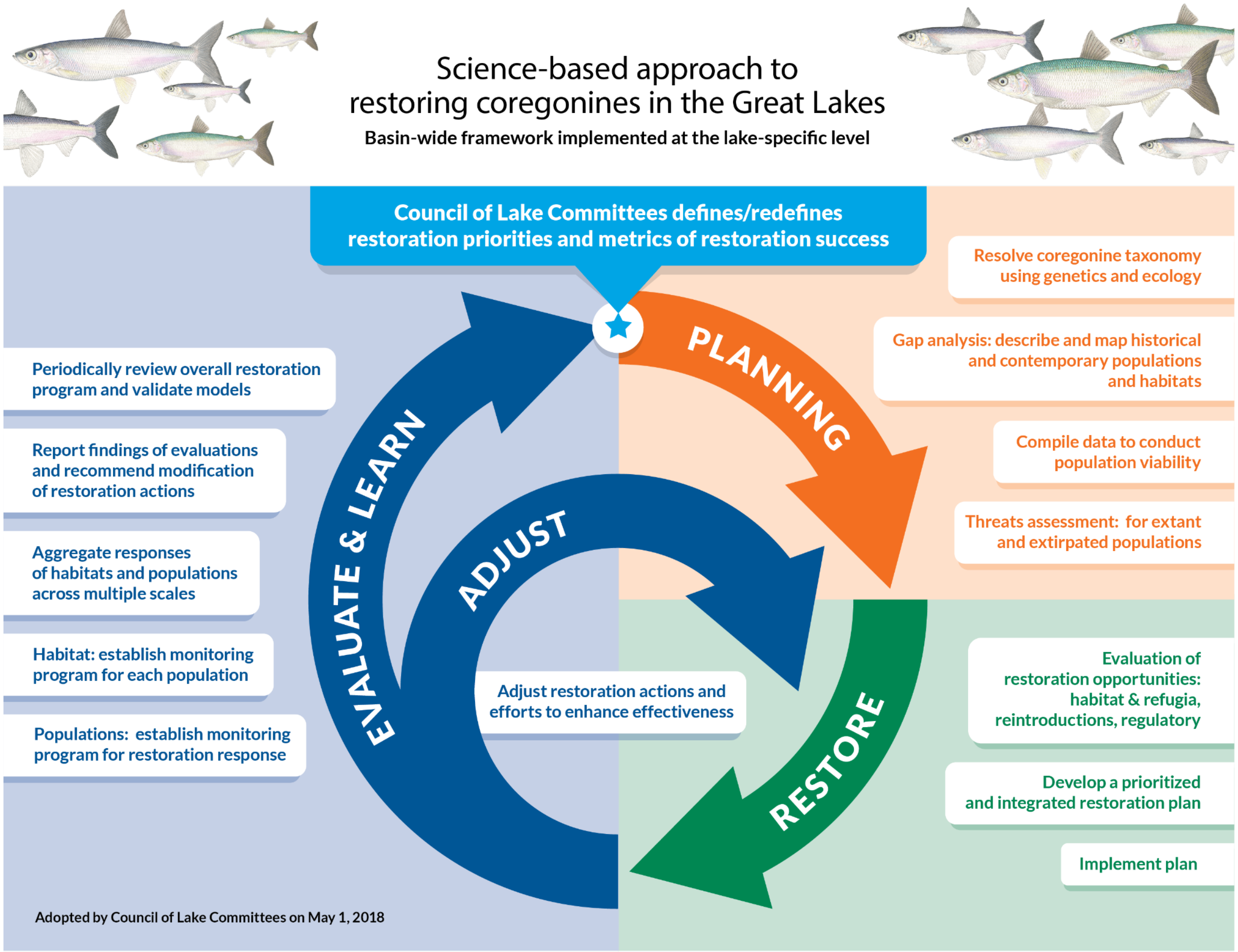


Figure 1: A knowledge and science-based approach to restoring coregonines in the Great Lakes. The Coregonine Restoration Framework is organized into three phases: Planning, Restore, and Evaluate, Learn and Adjust. Each phase encompasses individual tasks that together support coregonine restoration in the Great Lakes. Tasks and phases are not mutually exclusive; individual tasks may depend on outcomes from others within a phase and activities within each phase may be co-occurring. Credit: GLFC.

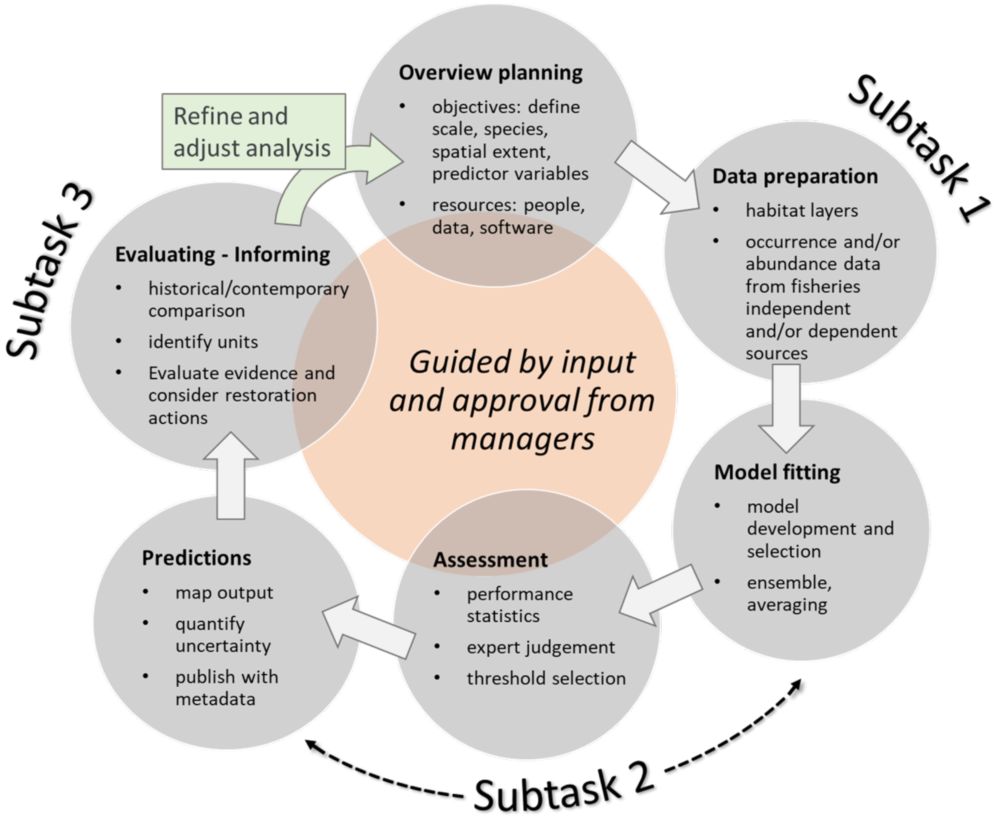


Figure 2. Steps involved in the process of the proposed Team 2 Gap Analysis. Subtasks I-III of the Gap Analysis methodology are labeled. Overlapping circles show where guidance from managers is needed. The proposed process involves 1) Overview planning (with guidance from managers), 2) Data preparation, 3) Model fitting, 4) Assessment (with guidance from managers), 5) Predictions, and 6) Evaluating - Informing (with guidance from managers). The analyses are then refined and adjusted as needed, based on manager's input.

Subtask I: Collect and Database spatial, temporal, historical and contemporary data on Coregonine spawning locations and occurrence.

A. Objectives

1. Gather, digitize and QA/QC historical and contemporary occurrence data.
2. Gather habitat layers (Great Lakes Aquatic Habitat Framework, historical substrate, threats/fishing pressure).
3. Make data accessible and usable for other tasks/subtasks.

B. Considerations

Data:

A foundational step of the Gap Analysis methodology is to collect, digitize, perform technical validation (quality assurance/control), and compile into a database the historical and contemporary sources and data related to coregonine occurrences in the Great Lakes. A wide range of sources exist that are related to coregonine occurrences in the Great Lakes and are diverse in format, data type, and accessibility. These sources must be tracked down, digitized, curated, subjected to quality assurance/control, and made machine readable. Many key sources, including collections of records (Appendix, Table A1) and manuscripts and/or reports (Table A2) have already been collected by this team, and are known to contain important coregonine occurrence data across the Great Lakes. Another key source, Indigenous Ecological Knowledge (IEK), also contains critical historical and contemporary information; however, IEK is not adequately documented in the Great Lakes. A relational database can be built specifically for housing and using historical fisheries data (Duarte et al. 2018). To be useful for mapping and describing historical and contemporary populations and habitats, historical occurrence data, at minimum, must contain reliable information on location, time, and taxonomy. Details of the event are also highly important (*i.e.*, sampling gear used, who made observations, the evidence that suggested fish were spawning, etc.).

Potential strategies, advantages, limitations, and biases associated with historical species occurrence data have been thoroughly reviewed (Tingley and Beissinger 2009; Pooley 2018). All historical data projects are subject to human biases and error (Duarte et al. 2018). One simple yet important way to reduce error at the onset of a data project of this scale is to begin extracting data from oldest sources first to better examine for duplication of data in later sources (Tingley and Beissinger 2009; Duarte et al. 2018). The quantity and reliability of historical occurrence data varies with 1) changes in fishing and sampling technology, 2) bias towards certain species and/or locations, 3) the quality and quantity of records, and 4) cultural filtering (Pooley 2018).

New technology (AUVs, LiDAR, video, more) is continuing to produce contemporary habitat data and identify new/novel spawning grounds, while recent and ongoing surveys and databases provide contemporary occurrence (all life stages) and confirm spawning ground data. Many data layers exist, are accessible to researchers and managers, and provide variable levels of information related to Great Lakes fishes, habitats, and other conditions (Appendix, Table A3). The most up-to-date layers can be queried, downloaded, and used to inform the Gap Analysis and other aspects of the CRF. An index of currently known data layers and databases that may inform the Gap Analysis for the CRF can be found in Table A3.

Finally, it is important to consider the fact that available habitat layers, whether they are historical or contemporary, are not always closely compatible with occurrence data. A data mismatch can occur spatially, temporally, or in spatial resolution. As an example, some data layers, like those that describe historical substrate, may be highly generalized and presented at low resolution for certain areas. These habitat layers will not always reflect local conditions, nor will they match the resolution of occurrence data that are georeferenced down to coordinates. While this is an inherent trade-off with these various types of data, some additional data can be added to existing layers to increase resolution for regions of interest. As an example, higher resolution historical substrate data can be extracted and compared from old U.S. Lake Survey maps (Woodford 1994)—work that is currently underway, being led by The Nature Conservancy.

Life history:

Information on coregonine spawning locations is commonly available in historical sources and is targeted in contemporary surveys (see examples in Tables A1 and A2). Spawning grounds are feasible target areas for conservation, restoration, and monitoring, since these physical locations are often described in detail in historical sources and through IEK, and can be efficiently mapped (Loftus 1980; Goodyear et al. 1982; Foster and Kennedy 1995; Fielder 2000; Duncan 2020; and many more, see Table A2). While spawning locations for multiple coregonine species have been summarized in several publications to date, our knowledge of spawning locations is incomplete. Other occurrence data across life history and habitats are less available in published literature (*i.e.*, nursery areas, migratory routes, feeding or overwintering grounds), yet can still be discovered and characterized from unpublished sources and IEK. For migratory species like coregonines, spawning grounds are one example of critical habitat (*i.e.*, repeatedly frequented and critical for population viability). Therefore, a spawning focus will support Task Team 3 methodology: population viability (Figure 1). Methodologies to address population viability and threats (Tasks 3 and 4) will incorporate life history parameters from additional stages (Figure 1).

Taxonomic accuracy:

Maintaining taxonomic accuracy is important to provide the best possible data from historical sources to inform the CRF. This can be difficult with coregonines in the Great Lakes – a wide range of vernacular is used to describe common names. Local and scientific names remain variable across regions and cultures, and over time. Extensive genetic and morphometric studies for coregonines in the Great Lakes are underway. For contemporary sources, taxonomic accuracy will be informed by Task 1 of the CRF: Resolve coregonine taxonomy using genetics and ecology (Figure 1). Any updates to coregonine taxonomy reported from Task Team 1 must be reflected and trackable in a database, as well as in all operations for extracting occurrence data from sources

Indigenous Ecological Knowledge (IEK):

Inclusion of Indigenous Peoples (First Nations and Tribes) and IEK is critical for the CRF to succeed. Indigenous communities have inhabited the Great Lakes region for thousands of years (*e.g.*, since time immemorial), creating deep connections between the people and the lands, waters, and native fishes within those waters. As the original caretakers and stewards of the waters, Indigenous Peoples have inherent and established rights with respect to Great Lakes fisheries. Such rights are upheld by federal and provincial/state courts, treaties, and reports like the Truth and Reconciliation Commission of Canada and United Nations Declaration on the Rights of Indigenous Peoples. Indigenous rights need to be recognized in decision-making processes.

Indigenous observations, stories, science, and experiential knowledge yields valuable quantitative and qualitative data (Berkes 2017) related to coregonine spawning and occurrence (Duncan 2020). Incorporating a combination of non-Indigenous and Indigenous sources can yield more diverse and representative data for spawning locations, seasons, habitats, and behaviors (Moller et al. 2004; Duncan 2020). Researchers must embrace Two-Eyed Seeing as a guiding framework for knowledge generation - learning to see the strengths and value of IEK through one lens and mainstream knowledge (*i.e.*, Western or positivist science) through the other, while using both together for everyone's benefit (Reid et al. 2021). IEK sources and data are often based on longer periods of observation, can incorporate larger sample sizes, and invite participation across researchers, harvesters, and other community members. Inclusion builds strong partnerships, leads to community consensus, and empowers Indigenous resource managers to continue governing natural resources throughout Tribal and First Nation territories.

A past unbalanced and unethical approach to research and fisheries management must be acknowledged. In addition to a general exclusion from decision-making tables, Indigenous communities were historically treated more as subjects and less as true partners in research. In the future, Indigenous communities must be treated as equitable partners and co-producers of products and knowledge. Researchers need to be transparent in proposed outcomes, must be inclusive, adhere to established and community-specific ethical standards (*e.g.*, OCAP: Ownership, Control, Access, Possession), and need to work to build rapport and respect with Indigenous communities (Duncan 2020; Reid et al. 2021).

C. Review of Approaches

A wide range of historical sources can inform management plans aimed at conservation and restoration of native species. Ferrer-Paris et al. (2014) used a combination of data sources with diverse methodological approaches, including contemporary bird surveys and historical distribution records and literature, to map occurrences for eight species of Amazon parrot, develop historical and contemporary species distribution models, and estimate potential changes in species distributions. Alexander et al. (2009) collected and processed Gulf of Maine cod fishing logbooks dating as far back as the 1860s to map fishing grounds and spawning areas, document snapshots of historical ecosystem requirements capable of sustaining abundant cod populations, and establish a historical-baseline population structure, all of which can be used to inform cod restoration. Even artifacts and texts dating as far back as the Middle Ages have been instrumental for evaluating and mapping historical fish populations and informing contemporary restoration projects (Duarte et al. 2018).

Digitizing and databasing records is imperative when using historical data, especially when information is from diverse sources with varying degrees of accuracy, specificity, and spatial resolution. Duarte et al. (2018) built a relational database, the first of its kind, to house historical distribution data for freshwater fishes in Portugal, that has made over one millennium of freshwater fish occurrence data publicly available. The database includes over 2,000 technically validated records across taxa at multiple spatial scales, and has already informed several scientific studies aimed at determining historical baseline conditions for populations and assessing historical occurrence limits tolerated by specific species (Duarte et al. 2018). A historical coregonine database, called CORHIST, was recently built by a team of scientists across agencies and will support the CRF (C. Brant, unpublished data). The purpose of CORHIST is to store technically validated, historical, coregonine spawning and occurrence data and make those data accessible and mappable. A detailed description of the CORHIST project can be found in the Appendix. Records added to CORHIST undergo a technical validation process where coregonine occurrence records are extracted from a source by a researcher, checked for duplication, georeferenced to the finest scale possible, checked for additional metadata related to credibility, and all associated information (*e.g.*, habitat, location, gear, taxonomy, and biological information) is entered into the database. All metadata and records in the database are then reviewed for accuracy by another researcher. An index of sources currently being processed for entry into CORHIST, and progress updates for each source, can be found in Table A1 (larger key collections) and Table A2 (manuscripts, various documents, and collections). Taxonomic acuity in CORHIST is maintained through the Integrated Taxonomic Information System (ITIS) naming convention (https://www.itis.gov/).

Contemporary data on coregonine spawning locations is available from multiple studies and databases around the Great Lakes basin that are currently in development and recently completed. These include studies of egg and larval surveys, surveys for river and lake spawning habitat, work to understand and convey IEK and multi-generation commercial fishing knowledge, and studies using new and advanced technology to image and categorize spawning habitats (Table A3). The Population Viability Analysis team (PVA) of the CRF recently assembled a "Survey of Surveys" which serves as an index of fish sampling databases across agencies around the Great Lakes. Paufve et al. (2022) recently conducted active and passive egg sampling surveys using pumps and egg mats at three historical Great Lakes spawning locations (embayments in Lakes Superior, Huron, and Ontario), providing critical contemporary data on spawning occurrence, as well as new insights regarding coregonine egg distribution and substrate selection. Their results challenge previous results that *C. artedi were* broadcast spawners lacking preference for certain substrates (Koelz 1929; Smith 1956; Goodyear et al. 1982), indicating that they may indeed show broad spawning habitat selection at the species level yet more specific substrate selection within certain populations or embayments. Research in coregonine egg distribution suggests that certain spawning substrates may be more important than others for certain populations, and this selectivity may be site and population specific.

Duncan (2020) conducted/reviewed and analyzed interviews and literature, as part of an investigation into the IEK of the Saugeen Ojibway Nation, to map coregonine spawning and fishing areas along the Saugeen (Bruce) Peninsula of Lake Huron and better understand coregonine migratory behavior. Duncan's investigations emphasize the importance of Two-Eyed Seeing (Bartlett et al. 2012). In addition to Duncan (2020), the Saugeen Ojibway Nation has been engaging with IEK and the Two-Eyed Seeing approach to coregonine research with lake whitefish. Through using both IEK (through interviews) and Western science, they are informing a wide array of projects including acoustic telemetry; occurrence, migration, and spawning mapping; spawning and critical habitat assessment with the application of underwater remotely operated vehicles; larval assessment; and addressing the stocking of fish in territorial waters. This approach and example serve as a model for other investigations and research into coregonines through Two-Eyed Seeing.

New technology is being implemented around the Great Lakes through multiple highly coordinated and binational projects (Table A3). As an example, work in Lake Michigan (Green Bay, WI) used hydroacoustic imaging along with egg and larval surveys to evaluate riverine spawning lake whitefish populations (Ransom et al. 2021). A large partnership is also working to use multibeam technology to produce high resolution maps of important reefs around Green Bay and Lake Michigan and egg and larval surveys are being conducted on most of these reefs (P. Esselman, personal communication).

D. Recommended Approach(es)

For Objective 1: *Gather, digitize, and QA/QC historical and contemporary occurrence data,* we recommend continuing the CORHIST project, which aims to research, process, and database historical information related to coregonine occurrence and spawning locations. Details of the CORHIST project, the process, progress, as well as the work still to be done can be found in the Appendix. Many of the current data sources in CORHIST were discovered with assistance from multiple Great Lakes agencies. We recommend asking Great Lakes agencies and partners to help track down more historical sources and data specific to spawning occurrence. The focus on spawning information could be revised in the future based on input from the Population Viability Team's analyses (Task Team 3, Figure 1; Rosenfeld 2006). Further, we recommend creating partnerships that empower Indigenous researchers and communities to lead investigations into IEK, working with partners to conduct and process oral histories and Indigenous catch data, to encourage co-production of products (maps, summaries, data) that will continue to inform the CRF. For contemporary data, we recommend continuing to follow scientific efforts underway that 1) collect oral history and interview data, 2) conduct spawning surveys in Great Lakes tributaries, connecting channels, and main lakes, 3) investigate the distribution, condition, and movement throughout early life-history, from egg deposition through larval and juvenile (fall of age-0 until ~age-3) life stages, 4) utilize new technology to map habitats and occurrence through imaging, eDNA, and other techniques, 5) continue and expand the telemetry network, and 6) conduct long term monitoring and sampling through collaborative efforts around the basin.

For Objective 2: *Gather existing habitat layers*, we recommend continuing to track and collect data layers produced around the basin by various agencies (Table A3). This will require regular inquiries of Great Lakes agencies and partners to ensure we continue to use the most up-to-date information available. We can facilitate gathering most up-to-date layers via regular communications/requests through GLFC Joint Strategic Plan Committees. We recommend a focus on finding existing data layers that are representative and comparable with different periods of spawning occurrence data. As an example, substantial data for coregonine spawning occurrence exists in the Koelz era (ca. 1917-1929) – matching these data with water temperature layers covering the same time period would be ideal. We further recommend promoting continued use and support of the Great Lakes Aquatic Habitat Framework (GLAHF), a standardized framework which enables modeling applications (Wang et al. 2015; https://www.glahf.org/).

For Objective 3: *Make data accessible and usable for other tasks/subtasks*, we recommend using the historical coregonine database (CORHIST, Appendix 2). The CORHIST database, along with all metadata and maps, will be subjected to a high degree of quality assurance and quality control and published using the FAIR standards (Findable, Accessible, Interoperable, and Reusable) through the U.S. Geological Survey's ScienceBase Data Release Tool (https://www.sciencebase.gov/catalog/) upon completion (anticipated completion of first data release is May 2024, with future updates possible).

Subtask II: Model, map, and evaluate species-habitat distributions

A. Objectives

1. Define historical vs. contemporary periods
2. Develop species distribution models for both historical and contemporary periods
3. Assess models and determine thresholds of habitat suitability
4. Map observations and model-based predictions of suitable habitat in both periods

B. Considerations

Historical vs. contemporary periods:

Historical and contemporary periods are defined in many ways, and these definitions are often context dependent (Tingley and Beissinger 2009). Periods are often distinguished by different data characteristics or sampling periods but can also reflect differences in environment or population demography. In defining these periods, we aim to a) capture a historical baseline or reference condition for populations and habitat, indicating where restoration may head, and b) understand the current distribution of populations and habitat (status of these populations being the focus of Team 3).

Environmental conditions have changed for coregonines over the last century due to invasive species introductions, climate change, and other anthropogenic factors, which may guide determination of historical versus contemporary periods in certain cases. While some of these disruptions are clearly defined events, others are temporally ambiguous, occurring to varying degrees since the early 19th century (Bogue 2000). Examples include sea lamprey invasion (1940s), alewife proliferating in the 1960s, phosphorus controls 1970s-80s, dreissenid mussel invasions (1980s and 1990s), salmon introductions (1960s-70s), pollution events (wood pulp ca. 1800s, mining ca. 1950s-70s, industrial pollution across several decades), deforestation, shoreline erosion and loss of wetlands and riparian areas, fluctuating and generally declining ice cover, significant channelization and/or habitat modification for shipping/commerce, or changes in fishing pressure (for example with limited entry for commercial fishing in Michigan enacted in 1970).

Data characteristics and availability (see *data quality and quantity* below) can vary across historical and contemporary periods. Historical biological data is often opportunistically collected or heterogeneous in resolution and accuracy. In contrast contemporary data are more likely to be derived from standardized long-term monitoring programs and spatially explicit digital databases. Across the five Great Lake basins, nutrients, plankton, and fish at various trophic levels have been monitored since the mid-1970s and data across all trophic levels are available since 1998 (Bunnell et al. 2014; Bulakova et al. 2018; Gorman 2019).

Data quality and quantity:

Mapping observations of fishes can allow us to identify changes or consistencies in population and habitat locations over time and space. However, observations can be biased by the distribution of sampling effort and poor delineation of where sampling actually occurred and whether species were not observed or there was no sampling. When data are sparse or sampling is unevenly distributed across locations (as can be the case with historical records), interpreting observations can lead to assuming that habitats and locations did not (or do not) support viable locations when they actually do (Type 1 error). Species distribution models (SDMs) describing the relationship between species occurrences and underlying environmental factors can be used to predict where species previously occurred and where they are likely to currently occur, even in unsampled locations. These models are built, tested, and validated with observational data which makes accurate data all that much more important.

When developing and interpreting species distribution models and maps there are several factors related to data quantity and quality that should be considered. Data type (presence-only, presence/absence, or abundance), amount, and completeness of spatial and temporal coverage can determine the suitability of modeling approaches. Gear used, total effort, and lack of statistical independence can create biases in interpretations and may violate underlying statistical assumptions of models.

Several types of data exist related to fish sampling events, all of which may be important for informing this methodology. Fishery-dependent data (commercial catch data) are often presence-only, meaning only catches are reported and not the absence of catches which are important in determining habitat suitability. Presence-only data are frequently used in modeling species distributions because they are the simplest form of occurrence data, are often widely available from museums and archives, and can provide powerful evidence for persistence over time for important habitats (Pennino et al. 2016). Caveats of this type of data are that the lack of information on non-detections leads to relatively poor habitat discrimination and can limit researchers' ability to determine range-shifts. Presence/absence data are produced when all detected (and non-detected) species are recorded. Fishery-independent data (scientific surveys) are most likely to include presence/absence data collected with standardized protocols and gear applied across sites and time periods to limit potential biases. By containing more information than presence-only data, presence/absence data can be more powerful for detecting changes in occupancy for specific areas (Pennino et al. 2016). These data, however, are often more limited in coverage than presence-only data. Further, presence/absence data do not discriminate non-detections from true absences, which can lead to overestimation of colonization events and underestimation of extinction events. Presence/absence data from surveys repeated across short time periods by a consistent observer or multiple independent observers can be used in occupancy models which account for imperfect detection and allow estimating the probability of detecting true presence or absence. Abundance is occasionally available in fishery-dependent and fishery-independent data; however, the interpretation of abundance relies upon understanding catchability (for a species with a gear type) and effort (e.g., number of nets and soak time). The lack of catchability and effort data, and the relative infrequency of abundance data, have often limited the use of abundance data in species distribution models (Howard et al. 2014). There are, however, recent and promising approaches for using abundance data such as monitoring biodiversity change (since abundance is more robust at detecting impacts on distribution than occurrence), as well as ecosystem function and services (such as monitoring interaction strength using abundance of species that interact) (Waldock et al. 2022).

Mapping standards:

Maps should be usable and accessible (*i.e.*, 508 compliant to ensure accessibility for users with disabilities, www.section508.gov). Maps should be consistently formatted, printable, and available through a web platform. A model example for mapping standards is GLAHF (https://www.glahf.org/).

C. Review of Approaches

Species distribution models (SDMs) have been used in numerous conservation and management applications ranging from identifying key habitat features, to mapping suitable habitat for species of conservation concern, to predicting shifts in species distributions under future climate change scenarios (Elith and Leathwick 2009, Guisan et al. 2017). Underlying each of these modeling approaches are a number of generalized steps (outlined in Guisan et al. 2017, Zurell et al. 2020 and modified and presented in Figure 2): 1) Overview planning, 2) Data preparation, 3) Model fitting, 4) Assessment, 5) Predictions. Subtask 1 entails data preparation. Model fitting, assessment and predictions all fit within Subtask 2. Subtask 3 represents an additional step, 6) Evaluating and informing, which along with overview planning and assessment, is guided by the input and approval of managers.

There are numerous species distribution modeling approaches with different but related statistical frameworks. We outline common types of models, their advantages and limitations related to data quality and quantity in Table A4. SDM approaches have been further reviewed by numerous authors who highlight important considerations and assumptions. (e.g., Peterson et al. 2011, Guisan et al. 2017). Most SDMs are either regression-based models (logistic regression, GLM/GLMM, GAM/GAMM) or models based in machine learning (CART, RF, BRT, Maxent). See Table A4, for model descriptions and comparisons. Less common, more data intensive, and recently developed approaches to modeling species distributions include occupancy models, joint species distribution models, hierarchical Bayesian models, and mechanistic models which can improve upon models based on species-habitat correlations (Carroll et al. 2010, Comte & Grenouillet 2013, Hayes et al. 2009, Minns et al. 1999, Pollock et al. 2014, Rougier et al. 2015, MacKenzie et al. 2018).

Since each modeling approach has its own strengths and weaknesses, no one approach is appropriate for all purposes and the choice of model can depend on the objectives of the study and the characteristics of available data (Elith and Leathwick 2009, Fukada et al. 2013). Several R packages (e.g., `biomod2` (Thuiller et al. 2009) and `dismo` (Hijmans et al. 2013) allow running and comparing multiple modeling algorithms. Best models can be selected by comparing overall predictive performance (correct identification of presences and absences). Alternatively, models with similar performance but differing predictions can be combined into an ‘ensemble’ which may be less sensitive to sampling bias and more transferable to different areas or time periods (Randin 2006). Ensembles can be useful for clarifying the main trend and overall variation across models (Guisan et al. 2017).

In fact, SDMs have been used in the Great Lakes Basin to identify fish habitat and its key features and the distributions of native and invasive species. Most of these SDMs have focused on areas of limited extent within lakes – a few studies have built basin-wide models. Beginning with regression-based approaches, generalized additive models (GAMs) of CPUE from the *R/V Fulmar* (1930-1932) dataset confirm historical accounts of depth segregation in the deepwater Cisco complex of Lake Michigan (Bunnell et al. 2012). More recently, Kao et al. (2020) used GAMs to describe historical (1930–1932) Cisco (*Coregonus artedi*) habitat use across seasons by combining the fishery-independent data collected by the *R/V Fulmar* with fishery-dependent data from the State of Michigan. Maps of the predicted catch of Cisco in Lake Michigan from these models highlight the importance of Green Bay and Grand Traverse Bay as fall spawning habitats and throughout the life cycle of Cisco. Finally, Brown et al. (2022) used generalized additive mixed models (GAMMs) to quantify the effects of regional climatic drivers (expressed as ice cover duration) and local habitat variables (distance to shore and depth) on the early life stage success (measured as larval counts) of Cisco and Whitefish in Lake Ontario.

Increasingly sophisticated machine learning-based models have also been used in Great Lakes applications. Schaefer (2022) fitted both classification and regression tree (CART) and random forest (RF) models to predict spawning habitat for Lake Whitefish and Cisco in Lake Erie and Lake Ontario. These models demonstrated that habitat suitability was related to combinations of ice cover, fetch, and substrate types (Schaefer 2019; Schaefer et al. 2022). RF models have also been used to classify coastal habitat use for 20 species of Great Lakes fishes by associating probability of occurrence with key habitat factors including chlorophyll concentrations, turbidity, wave height, and bottom slope (Kovalenko et al. 2018). Egly et al. (2019) used boosted regression trees to predict the potential distribution of Red Swamp Crayfish in the Great Lakes, identifying optimal habitat located within shallow bays and along the southern coastlines of lakes Michigan and Erie. Finally, G. Annis et al. (in prep.) used Maxent to predict spawning and nursery habitat for Cisco and Lake Whitefish across the Great Lakes.

D. Recommended Approach(es)

Our recommended approaches for defining periods, mapping, and modeling, are dependent upon whether the focus is basin-wide or lake/region specific and which coregonine species is being considered.

Objective 1: *Define historical vs. contemporary periods*

At a basin-wide scale, we propose a general definition of *historical* as up to the year1960, prior to the collapse of several coregonine forms according to available production data (Baldwin et al. 2009). Importantly, and as indicated in Figure 3 and Table 1, *historical* periods may in some cases be more appropriately defined at the lake-specific, and even regionally specific (within lakes), scales. Coregonine stock collapses can indicate changes in populations and habitat conditions, but the timing of these collapses is known to differ between species and locations. Using Cisco as an example, lake-specific collapse is often visible by the 1960s; however, major collapse occurred in Lake Erie in 1925 and gradual declines in Lake Ontario continued through the 1980s (Figure 3). When defining the end of the historical period at the lake-specific scale, or within lakes at regional scales, we recommend examining when and where a given species collapsed using available information. An example can be seen in Table 1 from Smith (1964), which shows declines in Lake Michigan across species, and regions, from the early-1930s to mid-1950s to early-1960s. A high degree of communication among partners within specific regions, across lakes, and across the basin, will be necessary to define these periods (along with historical research), and this communication can be facilitated with managers for Tribal, First Nations, state, and provincial agencies at meetings and conferences. Historical breakpoints could be considered lake, species, and/or embayment-specific and should be carefully derived so that model development incorporates data from a benchmark, or baseline, period which can guide restoration objectives. Because correlation-based species distribution models derive predictions based on the distribution of occurrences relative to habitat features, including data from post-collapse periods could bias historical distribution models. Data from such periods can, however, be informative for evaluating model predictions.

We recommend the basin-wide, *contemporary* period be defined as the year 2000 - present (Eshenroder et al. 2016), which accounts for dreissenid mussel invasions and ecosystem-scale change (Bunnell et al. 2014). Given the focus on habitat and population distribution mapping and modeling for the spawning stage and given the known impacts of dreissenid mussels on spawning habitats, juveniles, and forage (Hoyle et al. 1999), the contemporary period should begin after the spread of mussels through these habitats in the early 2000s. At the basin-wide scale, starting the contemporary period at the turn of the century is also appropriate when considering data quantity and quality as standardized contemporary survey data are also available from this period onward. We conclude with the suggestion that all definitions for historical and contemporary periods be agreed upon by partners, informed by expert knowledge, and reviewed and approved by managers prior to any further analyses.

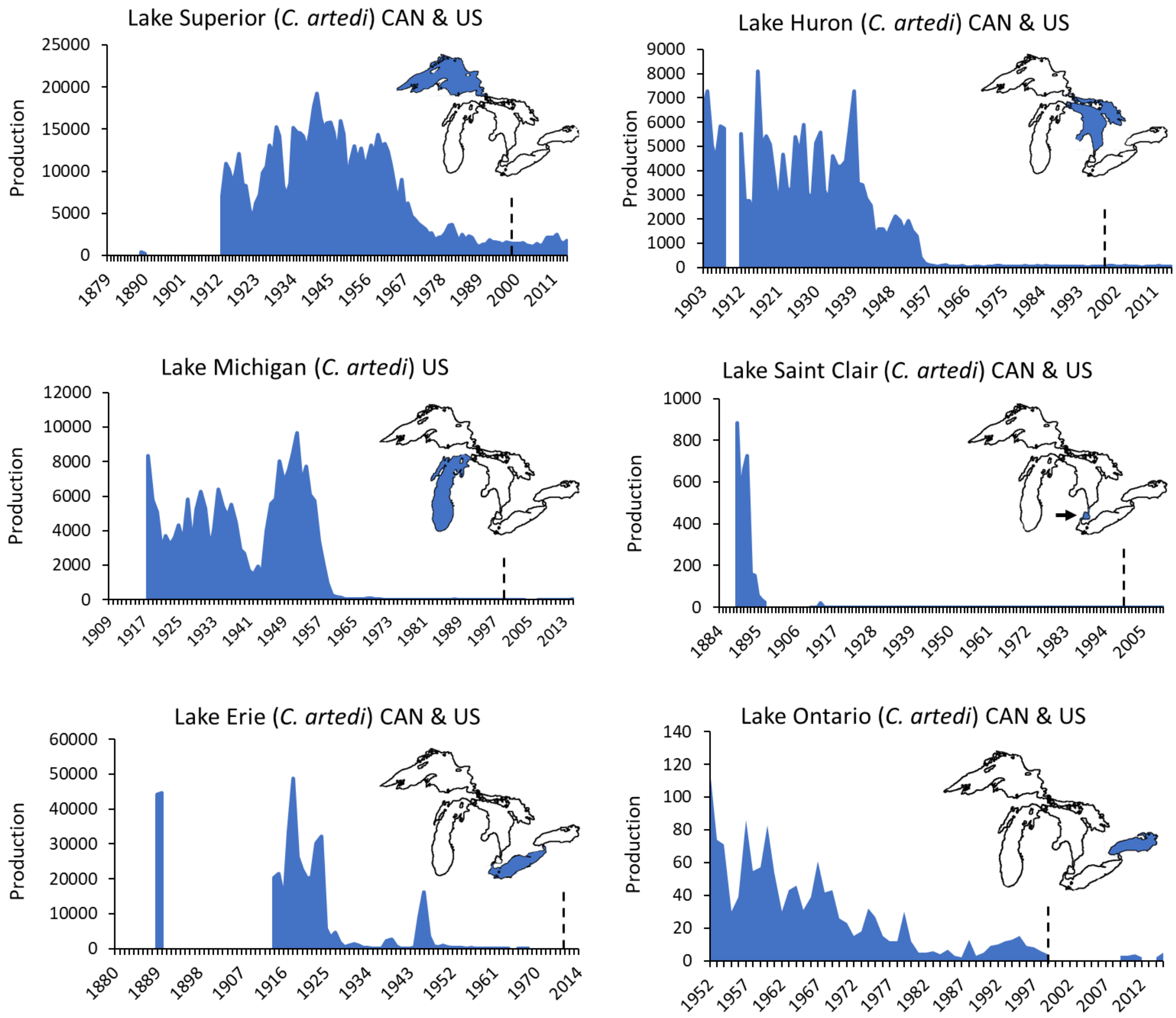
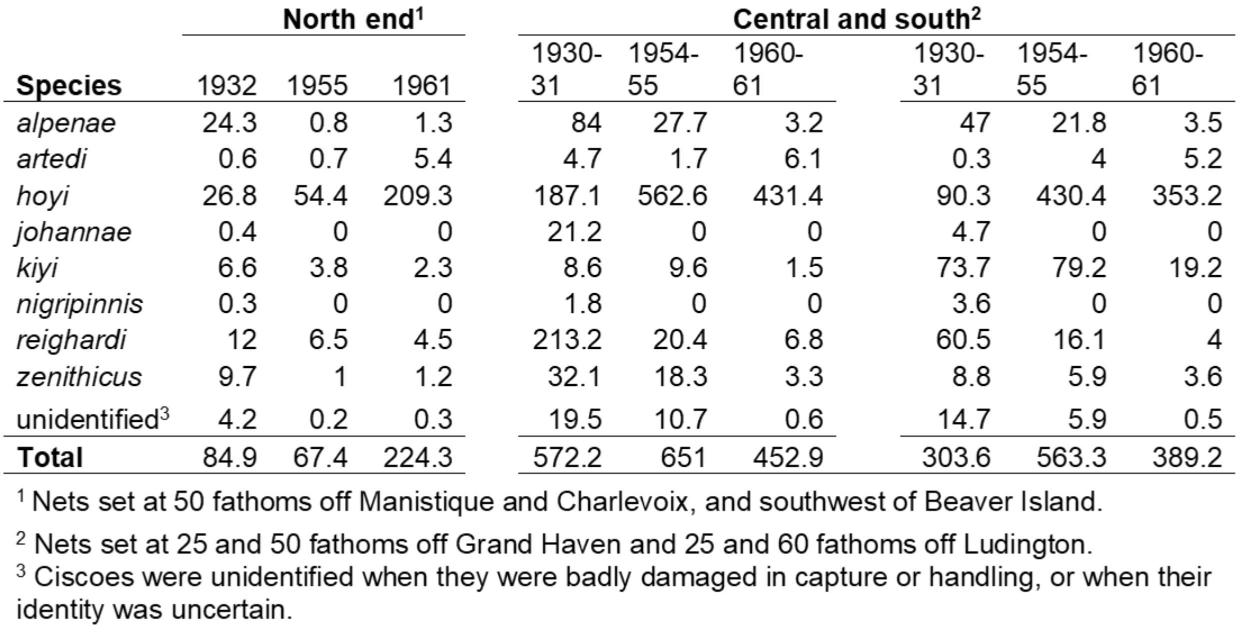


Figure 3. An example approach for determining historical versus contemporary breakpoints by lake and species (*C. artedi*) based on available commercial catch data. In general, breakpoints could be considered as follows: Lake Superior = 1960, Lake Huron = 1955, Lake Michigan = 1960, Lake Saint Clair = 1890s, Lake Erie = 1925, and Lake Ontario = 1980. Dotted reference lines mark the year 2000 (our proposed contemporary data start point). Production = annual catch in thousands of pounds (data source: Baldwin et al. 2009).

Table 1. A table transcribed from Smith (1964) shows the number of coregonines taken in Lake Michigan in identical gangs of gill nets across different periods and regions. Numbers for the north end are catch per 1530-ft gang, and central and south end are catch per 2550-ft gang. Smith (1964) represents an example of an approach for examining collapse breakpoints for multiple species and at multiple spatial and temporal scales.



	North end[1]			Central and south[2]					
Species	1932	1955	1961	1930-31	1954-55	1960-61	1930-31	1954-55	1960-61
alpenae	24.3	0.8	1.3	84	27.7	3.2	47	21.8	3.5
artedi	0.6	0.7	5.4	4.7	1.7	6.1	0.3	4	5.2
hoyi	26.8	54.4	209.3	187.1	562.6	431.4	90.3	430.4	353.2
johannae	0.4	0	0	21.2	0	0	4.7	0	0
kiyi	6.6	3.8	2.3	8.6	9.6	1.5	73.7	79.2	19.2
nigripinnis	0.3	0	0	1.8	0	0	3.6	0	0
reighardi	12	6.5	4.5	213.2	20.4	6.8	60.5	16.1	4
zenithicus	9.7	1	1.2	32.1	18.3	3.3	8.8	5.9	3.6
unidentified[3]	4.2	0.2	0.3	19.5	10.7	0.6	14.7	5.9	0.5
Total	84.9	67.4	224.3	572.2	651	452.9	303.6	563.3	389.2

[1] Nets set at 50 fathoms off Manistique and Charlevoix, and southwest of Beaver Island.

[2] Nets set at 25 and 50 fathoms off Grand Haven and 25 and 60 fathoms off Ludington.

[3] Ciscoes were unidentified when they were badly damaged in capture or handling, or when their identity was uncertain.

Objective 2: *Develop species distribution models for both historical and contemporary periods*
Model selection must be informed by the area and species under consideration, our understanding of the ecology of the focal species, availability of habitat and species occurrence data, and most importantly by objectives defined with input from managers. Basin-wide models for coregonines can highlight regions that would benefit from further work to address management needs with finer scale models and where high-quality data are available. Given that historical data on coregonine species can be patchy or relatively scarce and is commonly presence-only data, analysis at the basin-wide scale lends towards using Maxent modeling (Table A4). In cases where coregonine spawning biology has been observed to vary regionally, or between lakes, fitting lake-specific, or several regional models, can give better insight into habitat use than a single basin-wide model. As an example, Cisco spawning in Lake Ontario has recently been observed in highly protected embayments in contrast to some of the exposed shoreline spawning observed recently in the upper Great Lakes. In cases (locations, taxa, or periods) where presence-absence data, repeated sampling, or abundance data are available, other types of models or ensemble approaches comparing models are appropriate.

Objective 3: *Assess models and determine thresholds of habitat suitability*
Ultimately, model fitting will indirectly estimate the probability of species occurrence (or habitat suitability, on a zero to one scale) which can be mapped with gradients like heatmaps (e.g., Figure 4C). These heatmaps can be converted to distribution maps with a binary classification of areas with high probability of occurrence and suitable habitat presented and as ‘presence’ and low probability presented as ‘absences’ with thresholds fit to management applications. While a threshold of 0.5 is commonly used to separate high and low probability of occurrence, this threshold can lead to poor model performance depending on overall prevalence (the proportion

of presences). There are approaches to identify thresholds based on prevalence and on balancing sensitivity and specificity, but thresholds should be chosen based on the decision-making context (Guisan et al. 2017). Therefore, we recommend that a few thresholds be evaluated in coordination with managers to understand the implications of over or underpredicting habitat suitability. Ultimately, reclassification will be practical for identifying spatial units for restoration actions (see Subtask 3 below and assessment in Figure 4), however it should be recognized that converting continuous predictions to binary occupancy can degrade inference (Guillera-Arroita et al. 2015).

We recommend that SDM predictions be assessed with independent species occurrence data (data not used in model development) and with expert knowledge including review by managers and stakeholders. Models should be examined for calibration (extent to which conditional probability of presence is correctly predicted) and discrimination (distinguishing between occupied and unoccupied sites). Performance statistics can be threshold-dependent or threshold-independent (described in Guisan et al. 2017 and elsewhere). Threshold-dependent measures compare true positives, false positives, false negatives, and true negatives (e.g., true skill statistic (TSS), kappa, and correctly classified instances (CCI)). Threshold-independent measures compare predicted probability and presence/absence observations (e.g., mean squared error (MSE) and the commonly used area under the receiver operating characteristic curve (AUC)). Beyond statistical performance measures, we recommend that model outputs, both the key habitat features used in the model and mapped model outputs, be evaluated by managers, stakeholders, and others with expert knowledge of the species biology and distribution, geography, and habitat (Reside et al. 2019). This step can be crucial for identifying potential problems, missing data, or predictors, and refining future models to improve ecological realism. In this context, we further suggest that CRF modeling and mapping should be an iterative process informed by managers at several steps (Figure 2).

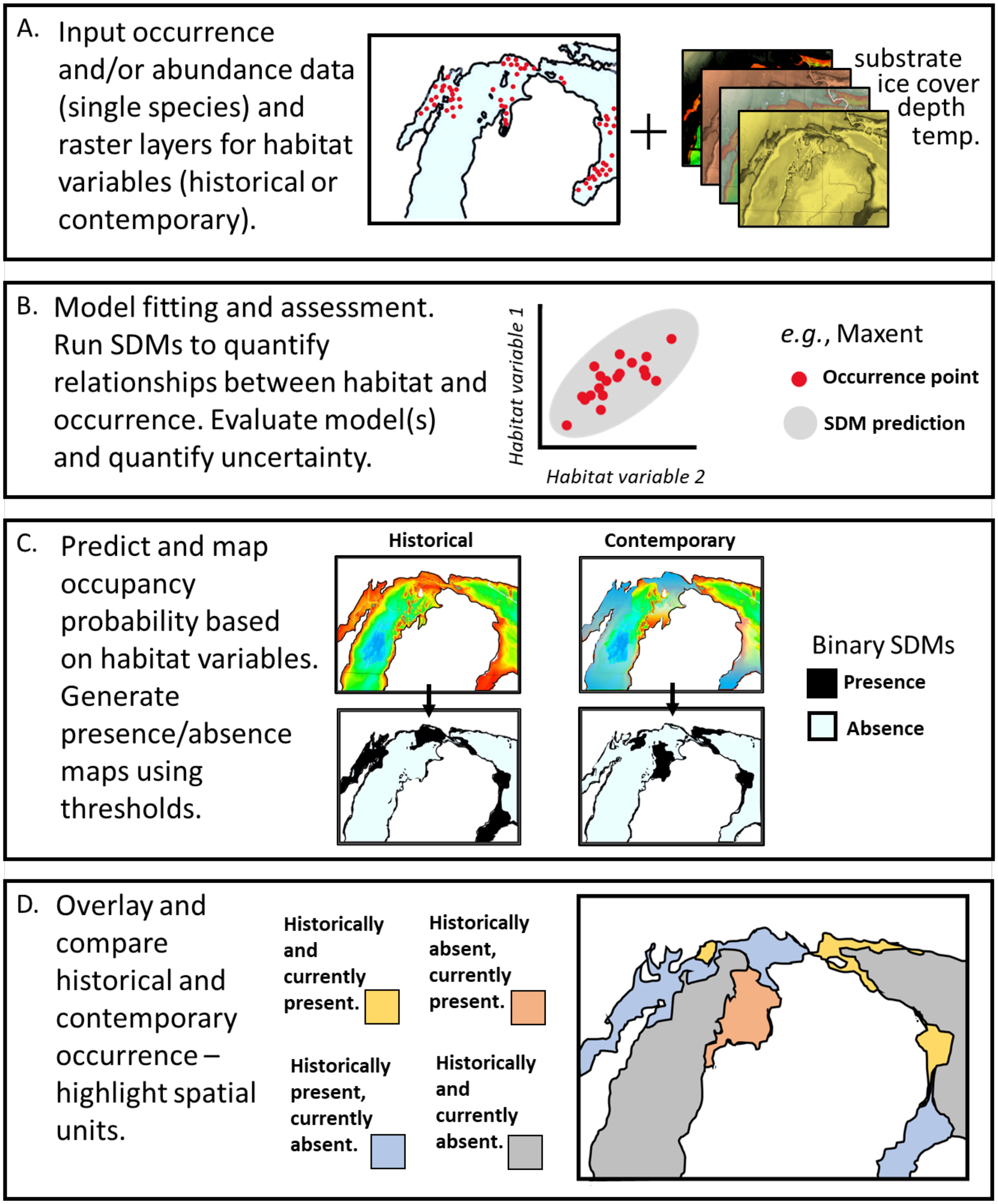


Figure 4. A summary of the mapping and modeling components of this methodology. A. Species occurrence and habitat data are gathered (Subtask I) to develop models. B. Models are built (Subtask II) to quantify the relationship between species occurrence and habitat variables. In C., predictions are mapped in the form of heat maps, where red represents a high probability of a species spawning. Probability of occurrence can also be reclassified as a binary (presence/absence) value using various thresholds (e.g., higher or lower than a 0.5 probability).

In D., historical and contemporary binary maps of predicted presence/absence (Subtask III) are overlaid to produce four conditions, which are described in our decision matrix (Table 2).

Objective 4: *Map observations and model-based predictions of suitable habitat in both periods*
The limited accessibility of model inputs and products has hampered the use of SDMs in management decision making (Sofaer et al. 2019). We strongly recommend that all modeling products be published and accessible to the Great Lakes fisheries community, as well as their underlying data and associated metadata depending on permissions. The GLAHF platform already houses SDMs for coastal habitats in a web-based mapping platform that allows visualizing underlying habitat drivers and predictions of species occurrence. We suggest future models built to support the Coregonine Restoration Framework be housed in GLAHF or a similarly publicly accessible platform, with archived geodatabases and metadata deposited in a repository such as ScienceBase.

III. ***Subtask III: "Gap Analysis" to compare historical and contemporary distributions***

A. Objectives

1. Compare historical and contemporary population and habitat distributions
2. Identify opportunities to inform coregonine conservation and restoration using a decision matrix.

B. Considerations

Alternative methods for comparing historical and contemporary distributions

While comparison of historical and contemporary population and habitat distributions can be accomplished by simply overlaying binary maps of species presence/absence in each period, overlap statistics (Fortin et al. 1996) also can be used to compare the spatial borders of population and habitat (spawning or other suitable habitats) distributions. The direct overlap statistic (O_D) compares how much of the boundaries overlap based on the number of boundary elements that are at the same location between periods. The minimum spatial distance (O_M) statistic quantifies whether there is a spatial lag between boundaries for the two periods. For example, if environmental conditions (e.g., water temperature) have spatially shifted through time thus affecting the boundary (i.e., location) of suitable spawning habitat in the contemporary period, the O_M statistic can be used to detect where along the boundary that shift may have occurred. The significance of these statistics is assessed using randomization tests (Fortin et al. 2005). The O_D and O_M approaches can be used to compare the overlap of individual polygons of populations and suitable habitat. Multiple polygons representing population and habitat can be compared between the two periods using six measures proposed by Sadahiro and Umemura (2001). These measures include the (1) generation: appearance of a polygon, (2) disappearance: loss of a polygon, (3) expansion: increase in area, (4) shrinkage: loss of area, (5) union: two polygons merging, and (6) division: a polygon splitting into two. Each of these can be measured using the number and area of each polygon and summarized for all polygons as the ‘event measure’ i.e., the synthesis of how many polygons were lost, modified, gained between the two time periods (Sadahiro and Umemura 2001; Fortin et al. 2005).

Alternatively, if thresholds are not applied and borders are not defined for populations and habitats, probabilities from the two time periods can be subtracted or compared as a ratio across locations. Here, probabilities could also be compared by the Structural Similarity (SSIM) index. The SSIM index uses a spatially-local window to calculate statistics based on local mean, variance, and covariance of locations or grids within the maps. A single summary statistic that provides an overall measure for comparison between periods can be calculated using the three components (Jones et al. 2016).

C. Review of Approaches

There have been many studies, of various taxa, that compared historical and contemporary species distributions and habitat use, although most of these studies focused on understanding species responses to climate change (Chen et al. 2011). Given limited historical and long-term data sources, however, more studies of the impacts of climate change have used SDMs to forecast species distributions under future climate scenarios (e.g., Van Zuiden et al. 2016). As an example of historical and contemporary comparisons, Ferrer-Paris et al. (2014) combined historical records with contemporary survey data to map eight species of parrot. SDMs were fitted to environmental covariates and presence-only data from museums and the literature to estimate the historical probability of occurrence for each species. The authors compared the species’ probability of occurrence from these historical SDMs with single-season occupancy models using recent survey data. Occupancy models enable time- and effort-dependent estimates of detectability but are data intensive and require repeated surveys of the same locations within short windows of time. Comparisons of models across time periods allowed the group to calculate and map the probability of change in the state of occurrence for each parrot species. For montane rodents, Pardi et al. (2020), likewise, compared SDMs developed with historical data and projected over current modern conditions with models developed using modern observations. They found changes in the importance of different climate predictors between models indicating changes in species-environment relationships. They also mapped changes in the probability of occurrence as a ratio of the two models. Parker et al. (2021) recently compared historical and contemporary SDMs of Headwater Catfish (*Ictalurus lupus*) built with Random Forests in order to assess conservation status. They used these models to 1) examine changes in the importance of various habitat variables, 2) map changes in habitat suitability (by subtracting suitability in the historical period from that in the contemporary period), and 3) identify sites with high habitat suitability and low levels of genetic introgression and hybridization with non-native catfishes as conservation targets. These applications demonstrate further potential uses of the SDMs that will be developed for the coregonine Gap Analysis.

Identification of spatial units or priority areas for conservation and restoration often involves the use of conservation-planning software. In recent years, two popular software packages have emerged: Marxan (Ball et al. 2009) and Zonation (Moilanen et al. 2009). Both software packages can help delineate conservation/restoration areas based on multiple metrics (*e.g.*, cost, accessibility, fishing pressure) and can incorporate SDMs for multiple species of interests. Delavenne et al. (2012) compared these widely used tools for delineating Marine Protected Areas in the eastern English Channel. The researchers found that the two are quite similar in their ability to set priority areas using primarily biological and social-economic data from the region, and both were strongly influenced by the cost metric. However, the team found that Zonation often produced results with greater connectivity (selecting for larger, more connected, patches), and outputs were more rectangular in shape. Marxan tended to produce more cost-efficient, patchy, areas with less regular boundaries. Overall, the two software packages produced similar results, but the authors note that future comparisons need to account for habitat which is essential for the completion of a fish life cycle, such as spawning grounds, and population dynamics – especially for exploited species (Delavenne et al. 2012). While not directly relevant to the Gap Analysis at this stage, in the future software packages like Marxan and Zonation may become important as the CRF continues into the implementation and refinement phases.

D. Recommended Approach(es)

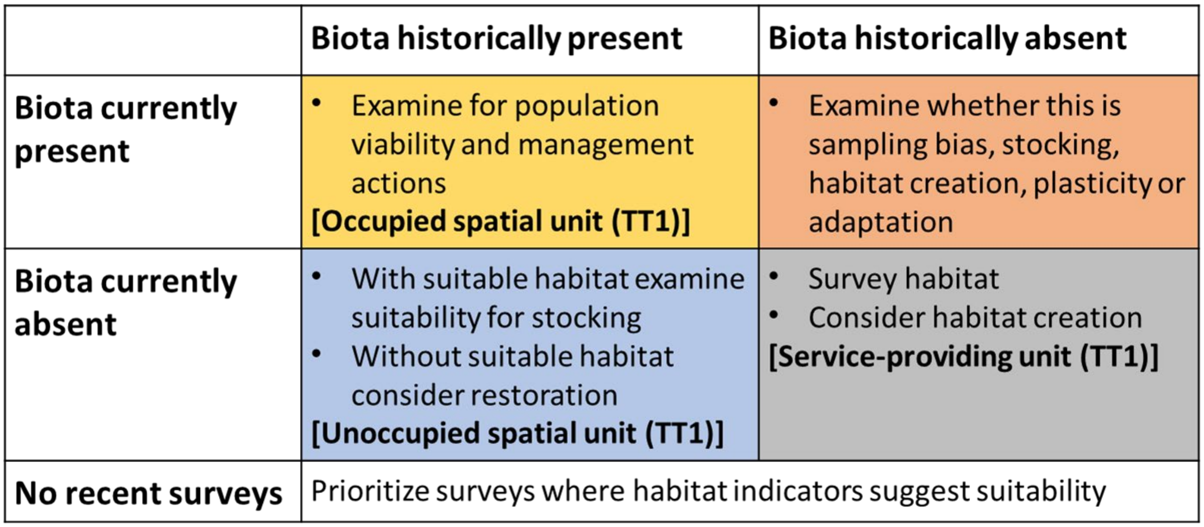
Objective 1: *Compare historical and contemporary population and habitat distributions*

While there are several approaches to comparing historical and contemporary distributions (reviewed above), many of these would be better suited for examining questions outside of the scope of this Gap Analysis. We recommend overlaying binary (presence/absence) maps of SDM predictions for each period (illustrated in Figure 4, panel D). Plotting available observations on these maps, with time periods indicated, would illustrate where there is evidence supporting classifications based on model predictions. Models may also predict species presence in areas with no recent surveys and we recommend these areas should be sampled to ground truth model predictions. Maps overlaying contemporary and historical distribution models and species observations also can be compared with underlying habitat data (e.g., substrate layers, bathymetry, temperature, etc.) in GIS or similar programs, allowing visualization of species-habitat relationships across the basin. Many spatial physical, chemical, and biological data are already available in the GLAHF explorer which is powered by ESRI software and freely accessible on the web. We recommend leveraging these tools.

Objective 2: *Identify opportunities to inform coregonine conservation and restoration using a decision matrix*

The four classes produced by overlaying the binary maps of predicted presence/absence can be arranged in a decision matrix (Table 2). In this matrix, an area in which a species was historically present, and still occurs (in yellow), represents what team 1 identifies as an *occupied spatial unit*. These populations could be further evaluated for population viability (PVA, Team 3) and threats assessment (Team 4). If a species was historically present but currently absent in an area (in blue), this represents what Team 1 defines as an *unoccupied unit*, which may be suitable for stocking or habitat restoration. If an area was never indicated as suitable spawning habitat, and a species was absent in both periods to our knowledge (gray), we recommend habitat surveys to evaluate the potential for habitat creation. Team 1 has identified these areas as potential *service-providing units*. Finally, there are areas where species were not known to historically occur, which now have coregonine spawning. These locations may have been historically undersampled or stocked, may be where habitat has been created, or may represent shifts in habitat use towards sub-optimal but currently available habitat. These locations also may represent cases of potential species plasticity or adaptation, as environmentally related plasticity in coregonines, including shifts in timing of spawning, has been documented in several studies (Svärdson 1950; Shields and Underhill 1993). Changes in species-habitat relationships can be identified by comparing the effect strengths of various habitat predictors across historical and contemporary distribution models. Further research is warranted on population viability under novel habitat conditions and whether such associations could inform stocking and restoration approaches.

Table 2: A decision matrix that can be used to identify management directions, opportunities for habitat restoration and stocking, and research based on comparisons of historical and contemporary presence and absence of spawning coregonines.



	Biota historically present	**Biota historically absent**
Biota currently present	• Examine for population viability and management actions **[Occupied spatial unit (TT1)]**	• Examine whether this is sampling bias, stocking, habitat creation, plasticity or adaptation
Biota currently absent	• With suitable habitat examine suitability for stocking • Without suitable habitat consider restoration **[Unoccupied spatial unit (TT1)]**	• Survey habitat • Consider habitat creation **[Service-providing unit (TT1)]**
No recent surveys	Prioritize surveys where habitat indicators suggest suitability	

References

Alexander, K.E., Leavenworth, W.B., Cournane, J., Cooper, A.B., Claesson, S., Brennan, S., Smith, G., Rains, L., Magness, K., Dunn, R, Law, T.K., 2009. Gulf of Maine cod in 1861: Historical analysis of fishery logbooks, with ecosystem implications. Fish and Fisheries, 10(4), 428-449.

Ball, I.R., Possingham, H.P. and Watts, M., 2009. Marxan and relatives: software for spatial conservation prioritisation. Spatial conservation prioritisation: Quantitative methods and computational tools, 14, pp.185-196.

Brown, T.A., Sethi, S.A., Rudstam, L.G., Holden, J.P., Connerton, M.J., Gorsky, D., Karboski, C.T., Chalupnicki, M.A., Sard, N.M., Roseman, E.F., Prindle, S.E., Sanderson, M.J., Evans, T.M., Cooper, A., Reinhart, D.J., Davis, C., Weidel, B.C., 2022. Contemporary spatial extent and environmental drivers of larval coregonine distributions across Lake Ontario. J. Great Lakes Res. 48, 359–370

Berkes, F. (2017). *Sacred Ecology* (4th ed.). Routledge. https://doi.org/10.4324/9781315114644

Bunnell, D.B., Eshenroder, R.L., Krause, A.E., Adams, J. V., 2012. Depth segregation of deepwater ciscoes (*coregonus* spp.) in lake michigan during 1930-1932 and range expansion of coregonus hoyi into deeper waters after the 1990s. Adv. Limnol. 63:3-24

Bunnell, D.B., Barbiero, R.P., Ludsin, S.A., Madenjian, C.P., Warren, G.J., Dolan, D.M., Brenden, T.O., Briland, R., Gorman, O.T., He, J.X., Johengen, T.H., 2014. Changing ecosystem dynamics in the Laurentian Great Lakes: bottom-up and top-down regulation. Bioscience, 64(1), 26-39.

Burlakova, L.E., Hinchey, E.K., Karatayev, A.Y., and Rudstam, L.G. 2018. U.S. EPA Great Lakes National Program Office monitoring of the Laurentian Great Lakes: insights from 40 years of data collection. Journal of Great Lakes Research 44(4): 535–538. doi:10.1016/j.jglr.2018.05.017.Carroll, C., Johnson, D.S., Dunk, J.R., Zielinski, W.J., 2010. Hierarchical Bayesian spatial models for multispecies conservation planning and monitoring. Conservation Biology, 24(6), 1538-1548.

Chen, I.C., Hill, J.K., Ohlemüller, R., Roy, D.B., Thomas, C.D., 2011. Rapid range shifts of species associated with high levels of climate warming. Science, 333(6045), 1024-1026.

Comte, L., Grenouillet, G., 2013. Species distribution modelling and imperfect detection: comparing occupancy versus consensus methods. Diversity and Distributions, 19(8), 996-1007.

Delavenne, J., Metcalfe, K., Smith, R.J., Vaz, S., Martin, C.S., Dupuis, L., Coppin, F., Carpentier, A., 2012. Systematic conservation planning in the eastern English Channel:

comparing the Marxan and Zonation decision-support tools. ICES Journal of Marine Science, 69(1), 75-83.

Duarte, G., Moreira, M., Branco, P., Da Costa, L., Ferreira, M.T., Segurado, P., 2018. One millennium of historical freshwater fish occurrence data for Portuguese rivers and streams. Scientific data, 5(1), 1-8.

Duncan, A.T., 2020. An investigation into the local and traditional knowledge of the Saugeen Ojibway Nation regarding the status of ciscoes (*Coregonus* spp.) in Lake Huron. Master's thesis, Lakehead University.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution and Systematics, 40(1), 677-697.

Eshenroder, R.L., Vecsei, Paul, Gorman, O.T., Yule, D.L., Pratt, T.C., Mandrak, N.E., Bunnell, D.B., Muir, A.M., 2016. Ciscoes (*Coregonus*, subgenus Leucichthys) of the Laurentian Great Lakes and Lake Nipigon: Great Lakes Fishery Commission Miscellaneous Publication 2016-01, accessed May 18, 2017, at http://www.glfc.org/pubs/misc/Ciscoes_of_the_Laurentian_Great_Lakes_and_Lake_Nipigon.pdf.

Ferrer-Paris, J.R., Sánchez-Mercado, A., Rodríguez-Clark, K.M., Rodríguez, J.P. and Rodríguez, G.A., 2014. Using limited data to detect changes in species distributions: Insights from Amazon parrots in Venezuela. Biological Conservation, 173, 133-143.

Fielder, D.G., 2000. Lake herring spawning grounds of the St. Marys River with some potential implications for early spring navigation. North American Journal of Fisheries Management, 20 (2), 552-561.

Fortin, M. J., Keitt, T. H., Maurer, B. A., Taper, M. L., Kaufman, D. M., Blackburn, T. M., 2005. Species' geographic ranges and distributional limits: pattern analysis and statistical issues. Oikos, 108(1), 7-17.

Fortin, M.-J., Drapeau, P., Jacquez, G. M., 1996. Quantification of the spatial co-occurrences of ecological boundaries. Oikos 77: 51-60

Foster, N.R. and Kennedy, G.W., 1995. Evaluation of Tawas artificial reef as fish spawning habitat. Methods of modifying habitat to benefit the Great Lakes ecosystem. Canadian Institute for Scientific and Technical Information, Occasional Papers, 1, 267-271.

Goodyear C.S., Edsall T.A., Dempsey D.M., Moss G.D., Polanski P.E. 1982. Atlas of the Spawning and Nursery Areas of Great Lakes Fishes. U. S. Fish and Wildlife Service, Washington, D.C., USA. FWS/OBS-82/52.

Gorman, O.T. 2019. Prey fish communities of the Laurentian Great Lakes: A cross-basin overview of status and trends based on bottom trawl surveys, 1978-2016. Aquatic Ecosystem Health & Management 22(3): 263–279. doi:10.1080/14634988.2019.1674012.

Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., McCarthy, M.A., Tingley, R. Wintle, B.A., 2015. Is my species distribution model fit for purpose? Matching data and models to applications. Global Ecology and Biogeography, 24(3), 276-292.

Guisan, A., Thuiller, W., Zimmermann, N. E., 2017. *Habitat suitability and distribution models: with applications in R*. Cambridge University Press.

Hayes, D., Jones, M., Lester, N., Chu, C., Doka, S., Netto, J., Stockwell, J., Thompson, B., Minns, C.K., Shuter, B., Collins, N., 2009. Linking fish population dynamics to habitat conditions: insights from the application of a process-oriented approach to several Great Lakes species. Reviews in Fish Biology and Fisheries, 19(3), 295-312.

Howard, C., Stephens, P.A., Pearce-Higgins, J.W., Gregory, R.D., Willis, S.G., 2014. Improving species distribution models: the value of data on abundance. Methods in Ecology and Evolution, 5(6), 506-513.

Hoyle, J.A., Schaner, T., Casselman, J.M., Dermott, R., 1999. Changes in lake whitefish (*Coregonus clupeaformis*) stocks in eastern Lake Ontario following Dreissena mussel invasion. Great Lakes Res. Rev, 4, 5-10.

Jones, E. L., Rendell, L., Pirotta, E., Long, J.A., 2016. Novel application of a quantitative spatial comparison tool to species distribution data. Ecological Indicators, 70, 67-76.

Kao, Y.-C., Bunnell, D.B., Eshenroder, R.L., Murray, D.N., 2020. Describing historical habitat use of a native fish—cisco (*Coregonus artedi*)—in Lake Michigan between 1930 and 1932. PLoS ONE 15(4): e0231420.

Koelz, W., 1929. Coregonid fishes of the Great Lakes. US Government Printing Office, Washington, D.C.

Kovalenko, K.E., Johnson, L.B., Riseng, C.M., Cooper, M.J., Johnson, K., Mason, L.A., McKenna Jr, J.E., Sparks-Jackson, B.L., Uzarski, D.G., 2018. Great Lakes coastal fish habitat classification and assessment. Journal of Great Lakes Research, 44(5), 1100-1109.

Loftus, D.H., 1980. Interviews with Lake Huron commercial fishermen. Lake Huron Fisheries Assessment Unit, Ontario Ministry of Natural Resources, Owen Sound, ON.

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., Hines, J.E. 2018. Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence. 2nd ed. Academic Press, San Diego, Calif.

Minns, C.K., Doka, S.E., Bakelaar, C.N., Brunette, P.C., Schertzer, W.M., 1999. Identifying Habitats essential for pike *Esox lucius* L. in the Long Point region of Lake Erie: a suitable supply approach. In Fish habitat: essential fish habitat and rehabilitation. American Fisheries Society, Symposium (Vol. 22, 363-382).

Moilanen, A., Kujala, H., and Leathwick, J. R. 2009. The Zonation framework and software for conservation prioritization. In Spatial Conservation Prioritization: Quantitative Methods and Computational Tools, pp. 196 –210. Ed. by A. Moilanen, K. A. Wilson, and H. P. Possingham. Oxford University Press, Oxford, UK. 320 pp.

Moller, H., Berkes, F., Lyver, P.O.B., Kislalioglu, M., 2004. Combining science and Traditional Ecological Knowledge: Monitoring populations for co-management. Ecology and Society 9(3). [online] URL: http://www.ecologyandsociety.org/vol9/iss3/art2

Pardi, M.I., Terry, R.C., Rickart, E.A., Rowe, R.J., 2020. Testing climate tracking of montane rodent distributions over the past century within the Great Basin ecoregion. Global Ecology and Conservation, 24, p.e01238.

Parker, S.D., Perkin, J.S., Bean, M.G., Lutz-Carrillo, D., Acre, M.R., 2021. Temporal distribution modelling reveals upstream habitat drying and downstream non-native introgression are squeezing out an imperiled headwater fish. Diversity and Distributions, 27(3), 533-551.

Paufve, M.R., Sethi, S.A., Weidel, B.C., Lantry, B.F., Yule, D.L., Rudstam, L.G., Jonas, J.L., Berglund, E., Connerton, M.J., Gorsky, D., Herbert, M.E., 2022. Diversity in spawning habitat use among Great Lakes Cisco populations. Ecology of Freshwater Fish, 31(2), 379-388.

Pennino, M.G., Conesa, D., Lopez-Quilez, A., Munoz, F., Fernández, A., Bellido, J.M., 2016. Fishery-dependent and-independent data lead to consistent estimations of essential habitats. ICES Journal of Marine Science, 73(9), 2302-2310.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araújo, M.B. (2011). *Ecological Niches and Geographic Distributions*. Princeton University Press.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A., McCarthy, M.A., 2014. Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). Methods in Ecology and Evolution, 5(5), 397-406.

Pooley, S., 2018. Descent with modification: Critical use of historical evidence for conservation. Conservation Letters. 11:e12437. https://doi.org/10.1111/conl.12437

Reid, A.J., Eckert, L.E., Lane, J.F., Young, N., Hinch, S.G., Darimont, C.T., Cooke, S.J., Ban, N.C., Marshall, A., 2021. "Two-Eyed Seeing": An Indigenous framework to transform fisheries research and management. Fish and Fisheries, 22(2), 243-261.

Reside, A.E., Critchell, K., Crayn, D.M., Goosem, M., Goosem, S., Hoskin, C.J., Sydes, T., Vanderduys, E.P. and Pressey, R.L., 2019. Beyond the model: expert knowledge improves predictions of species' fates under climate change. Ecological Applications, 29(1), p.e01824.

Rosenfeld, J.S., Hatfield, T., 2006. Information needs for assessing critical habitat of freshwater fish. Canadian Journal of Fisheries and Aquatic Sciences, 63(3), 683-698.

Rougier, T., Lassalle, G., Drouineau, H., Dumoulin, N., Faure, T., Deffuant, G., Rochard, E., Lambert, P., 2015. The combined use of correlative and mechanistic species distribution models benefits low conservation status species. PLoS One, 10(10), p.e0139194.

Sadahiro, Y., Umemura, M., 2001. A computational approach for the analysis of changes in polygon distributions. Journal of Geographical Systems, 3(2), 137-154.

Schaefer, H. (2019) Predicting spawning habitat for Lake Whitefish (*Coregonus clupeaformis*) and Cisco (*Coregonus artedi*) in the Lake Erie and Lake Ontario regions using classification and regression tree (CART) and random forest models. Michigan State University. Masters Thesis.

Schaefer, H.M., Honsey, A.E., Bunnell, D.B., Weidel, B.C., DeBruyne, R., Diana, J.S., Gorksy, D. and Roseman, E.F., 2022. Predicting physical and geomorphic habitat associated with historical lake whitefish and cisco spawning locations in Lakes Erie and Ontario. Journal of Great Lakes Research, 48(6), pp.1636-1646.

Sofaer, H.R., Jarnevich, C.S., Pearse, I.S., Smyth, R.L., Auer, S., Cook, G.L., Edwards Jr, T.C., Guala, G.F., Howard, T.G., Morisette, J.T., Hamilton, H., 2019. Development and delivery of species distribution models to inform decision-making. BioScience, 69(7), 544-557.

Tingley, M.W. and Beissinger, S.R., 2009. Detecting range shifts from historical species occurrences: new perspectives on old data. Trends in Ecology & Evolution, 24(11), 625-633.

Van Zuiden, T.M., Chen, M.M., Stefanoff, S., Lopez, L., Sharma, S., 2016. Projected impacts of climate change on three freshwater fishes and potential novel competitive interactions. Diversity and Distributions, 22(5), 603-614.

Wang, L., Riseng, C.M., Mason, L.A., Wehrly, K.E., Rutherford, E.S., McKenna Jr, J.E., Castiglione, C., Johnson, L.B., Infante, D.M., Sowa, S., Robertson, M., 2015. A spatial classification and database for management, research, and policy making: The Great Lakes aquatic habitat framework. Journal of Great Lakes Research, 41(2), 584-596.

Waldock, C., Stuart-Smith, R.D., Albouy, C., Cheung, W.W., Edgar, G.J., Mouillot, D., Tjiputra, J., Pellissier, L., 2022. A quantitative review of abundance-based species distribution models. Ecography, 2022(1).

Woodford, A.M., 1994. Charting the inland seas: A history of the US Lake Survey. Wayne State University Press.

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., … Merow, C. (2020). A standard protocol for reporting species distribution models. Ecography, 43(9), 1261–1277.