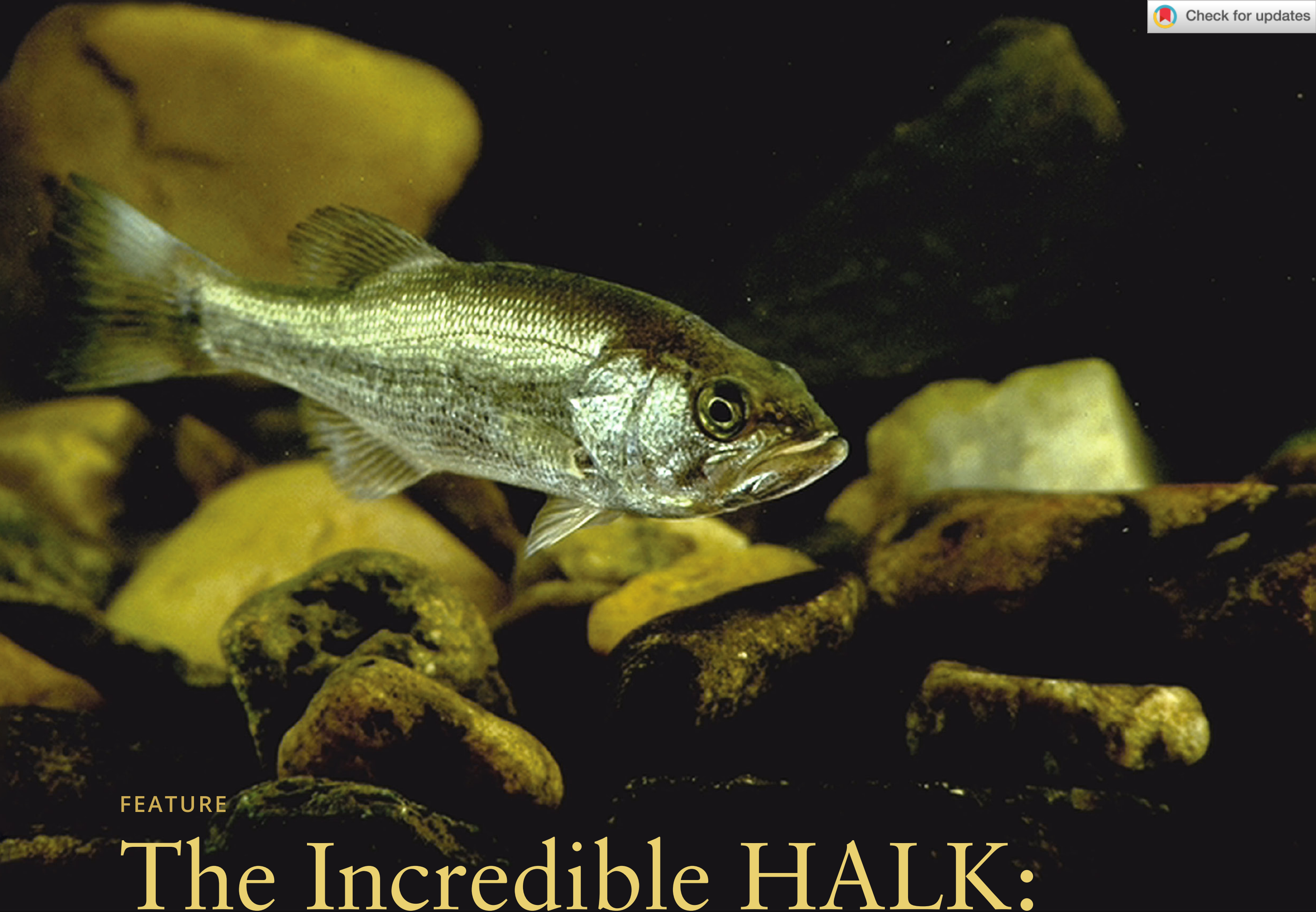

FEATURE

The Incredible HALK: Borrowing Data for Age Assignment

Paul N. Frater | Center for Limnology, University of Wisconsin-Madison, Madison, Wisconsin | Wisconsin Department of Natural Resources, Bureau of Fisheries Management, Madison, Wisconsin. E-mail: paul.frater@wisconsin.gov

Zachary S. Feiner | Center for Limnology, University of Wisconsin-Madison, Madison, Wisconsin | Wisconsin Department of Natural Resources, Office of Applied Science, Madison, Wisconsin

Gretchen J.A. Hansen | Department of Fisheries, Wildlife, and Conservation Biology, University of Minnesota–Twin Cities, St. Paul, Minnesota

Daniel A. Isermann | U.S. Geological Survey, Wisconsin Cooperative Fishery Research Unit, University of Wisconsin-Stevens Point, Stevens Point, Wisconsin

Alexander W. Latzka | Wisconsin Department of Natural Resources, Bureau of Fisheries Management, Madison, Wisconsin

Olaf P. Jensen | Center for Limnology, University of Wisconsin-Madison, Madison, Wisconsin

Largemouth Bass *Micropterus nigricans*. Photo credit: Ken Hammond, U.S. Department of Agriculture.

DOI: 10.1002/fsh.11019

Understanding age and growth are important for fisheries science and management; however, age data are not routinely collected for many populations. We propose and test a method of borrowing age–length data across increasingly broader spatiotemporal levels to create a hierarchical age–length key (HALK). We assessed this method by comparing growth and mortality metrics to those estimated from lake–year age–length keys ages using seven common freshwater fish species across the upper Midwestern United States. Levels used for data borrowing began most specifically by borrowing within lake across time and increased in breadth to include data within the Hydrologic Unit Code (HUC) 10 watershed, HUC8 watershed, Level III Ecoregion, and finally a species-wide data ALK using all available data with our study for a species. Median deviation in mean length of age-3 fish was within 1 cm for the most specific HALK levels, and median deviation in total annual mortality was close to 0 for most species when borrowing occurred within HUC10 and HUC8 watersheds. Percent error in growth curves increased with data borrowing, but plateaued—or even decreased—for some species when data borrowing expanded across spatial levels. We present the HALK as a method for gaining age information about a fishery when age data are unavailable.

The most valuable information obtained from sampled catch, at least for temperate waters, is age.
–Ray Hilborn and Carl J. Walters, 1992

INTRODUCTION

Age information is a crucial component for many aspects of fisheries science and management (Hilborn and Walters 1992; Kerns and Lombardi-Carlson 2017; Frater 2020). Fish ages are often used for estimating growth (Francis 2016; Lorenzen et al. 2016), mortality (Lee et al. 2011; Sippel et al. 2017), and year–class strength (Quist 2007). Age data are also important in many stock assessment models (Ono et al. 2015) and for estimating recruitment parameters (Magnusson and Hilborn 2007). However, age data are time-intensive and costly to collect, and obtaining the most precise and accurate ages for many species requires fish sacrifice (e.g., removal of internal structures such as otoliths or cleithra; Maceina et al. 2007). In contrast, measuring fish lengths is low-effort and less expensive, and, as a result, length data are more widely available (Quinn and Deriso 1999). Length data are valuable in their own right by providing information on population size structure (Froese 2004; Mildenberger et al. 2017) and long-term changes in body size resulting from environmental change (Oke et al. 2020; Solokas et al. 2023). Additionally, ages can be estimated from length and used to inform subsequent calculations (e.g., mortality). However, uncertainty in age estimation generated from fish lengths is propagated through subsequent calculations. As such, fisheries managers and scientists are faced with a trade-off—whether it is worth the additional resources to collect age data to inform more precise estimates, or whether those resources could be better used elsewhere in managing or monitoring fisheries (Hansen and Jones 2008).

Using a length-stratified subsample, an age–length key (ALK) can be created to assign ages to fish in the entire sample based on the observed proportion of fish at each age within each length strata or group (Isermann and Knight 2005). This nearly 100-year-old technique (Fridriksson 1934) is still commonly used in fisheries (Ailloud and Hoenig 2019; Ogle et al. 2023) and is based on the proportion of fish at age a as:

$$p_a = \sum \left(k_i p_{a \vee i}\right) \tag{1}$$

where i represents the discrete length bins from 1 to K, $p_{a|i}$ is the proportion of being age a given length bin i, and p_a is the total proportion in each age group.

Many resource management agencies collect length (as well as some age) data during standardized surveys (e.g., WDNR 2013; MNDNR 2017). However, due to logistical constraints, age data are often limited or unavailable for a specific water body and/or year. Paired age–length data may be available for other years and/or nearby water bodies, and age and length data could be borrowed from these to generate ALKs. This is likely to result in a loss of precision in age assignment and associated metrics calculated from age, but if precision loss is within acceptable tolerances, this approach would reduce resources spent on individual surveys to collect age data. The goal for this study was to develop a method to borrow data for age assignment that uses hierarchically nested ALKs to assign age to fish when age data are limited or nonexistent for a particular water body or time period and to evaluate the loss of precision in commonly used life history metrics calculated when age data are borrowed across time or space.

We present a method that borrows paired age–length data across multiple levels to create ALKs and assess the method using age–length data from seven freshwater fish species in over 5,000 lakes in the upper Midwestern United States. This method uses a lake–year ALK to assign age to fish when possible, but creates aggregate ALKs using data borrowed from increasingly broader levels, which are typically nested hierarchically. An aggregate ALK is created at each level where sufficient data are available, resulting in a series of hierarchical ALKs (HALKs). In this context, the most specific available ALK can be used to assign age to fish where age data are insufficient or lacking. This method extends the standard ALK from equation 1 by creating an aggregate ALK at each level, such as:

$$p_{a,L} = \sum \left(k_{i,L} p_{a,L \vee i}\right) \tag{2}$$

where the terms are all the same as in equation 1, but a subscript L is added to represent the level at which age–length data are aggregated. The HALK is the collection of these aggregated ALKs.

We introduce the R package *halk* that allows users to create their own HALKs and use them to assign age to length data. In this study, we use the package to simulate HALKs across levels for seven species common to the upper Midwestern United States, and we compare estimates of growth, mortality, and mean length at age to those of the lake–year ALK to assess changes in error of these estimates across HALK levels. Additionally, we developed an R Shiny application (https://bit.ly/3RuauJ6) that can be used to assign ages on length data using the HALKs from this study (Frater 2023). This application uses the HALKs developed in the data set used in this paper to assign age based on length using the most specific HALK level available.

METHODS

Building a Hierarchical Age–Length Key

A HALK is created using paired age–length data obtained from samples collected from different water bodies in different years. When age data sufficient for creating a lake–year ALK exist for a specific water body in a specific year (see below for data sufficiency criterion), no borrowing of ages need occur. If year-specific data within a water body are insufficient, the hierarchy of borrowing data to assign ages based on lengths can follow a progression defined by the user. For example, the progression could start by using age–length data collected in the same water body but different year(s) from the same water body, and then expand to using age–length data from other water bodies in the same watershed, eventually progressing to age assignment at the broadest level, which would be an ALK constructed using the entire provided age–length data set for a species (Figure 1). If a user were to specify levels of watershed, water body, and year, and provided age–length data for multiple water bodies across multiple years, then an ALK would be created for each unique water body *x* year combination (i.e., the lake–year ALK), each water body (i.e., pooling age–length data across years for a specific water body), each watershed (i.e., pooling age–length data for all water bodies in a watershed), and lastly, an ALK would be constructed using all available data for the species. Data requirements that determine if an ALK is created within each level—such as minimum number of age groups, samples per age group, or total samples—are defined by the user. At each hierarchy, age–length data are combined across the entire level to create an aggregate ALK. Fish that are measured for length—but not aged—are subsequently assigned an age using the ALK with the highest level of specificity.

The HALK procedure is automated in the R package *halk* using the function make_halk, where users provide age–length data and levels to be used in the age assignment hierarchy. This flexible user input maximizes the function's applicability to a wide range of situations. A fully reproducible working example for HALK creation and age assignment can be found in a vignette within the *halk* R package.

Application of the HALK to Sport Fish Data from Upper Midwestern Lakes

To demonstrate the application of the HALK method, we used fisheries monitoring data collected by state agencies in the upper Midwestern United States (Illinois, Indiana, Iowa, Michigan, Minnesota, South Dakota, and Wisconsin). This data set contained more than 23 million fish records (including 2.2 million age–length records) from standardized fishery

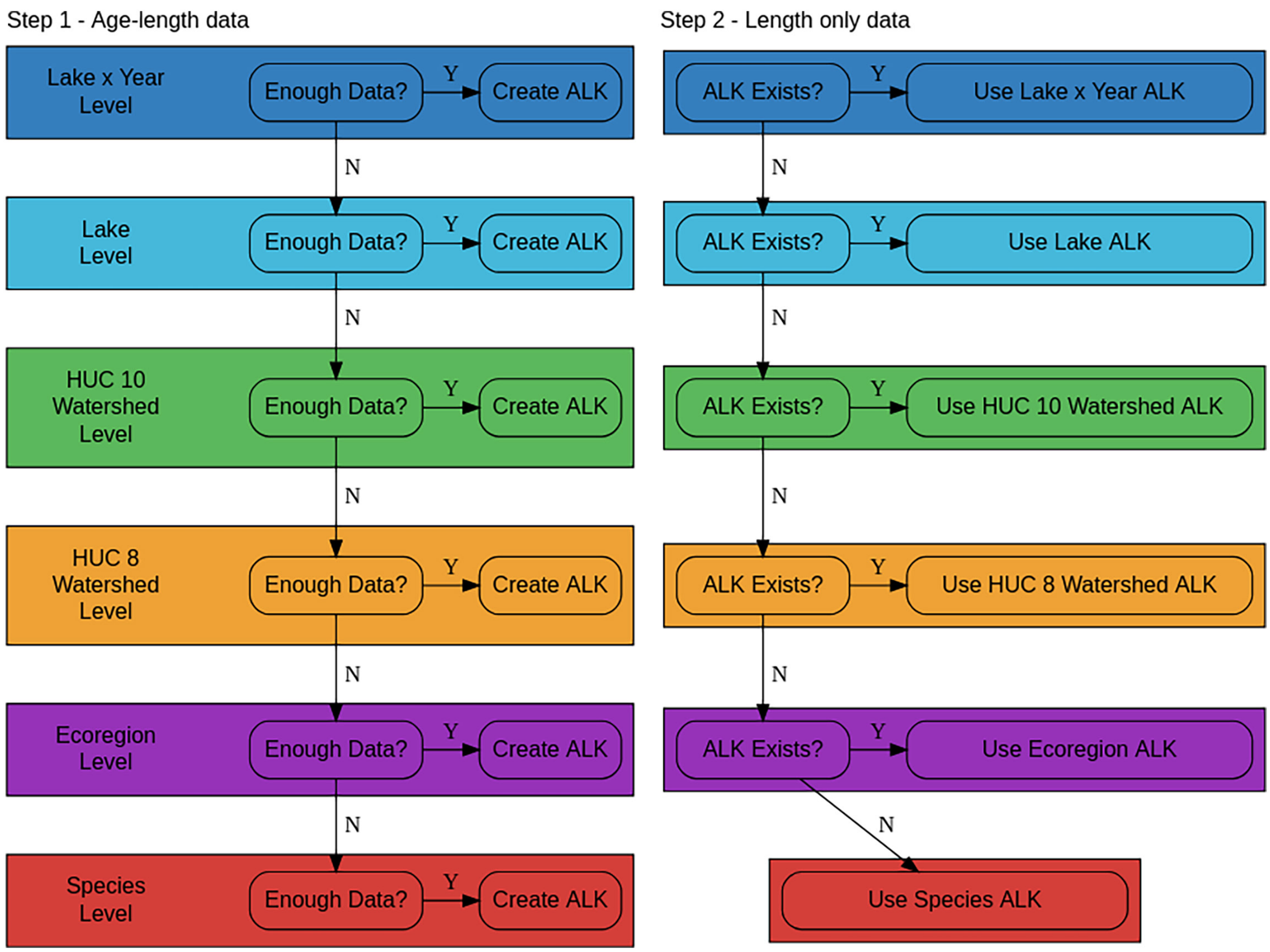

Figure 1. A schematic representation of the hierarchical age–length key (HALK) method. Levels of borrowing ages are specified by a user, and the number of samples of paired age–length data are checked at each level to determine if sufficient data exist (as specified by user). If sufficient data are available, an ALK is created; once a level is complete the process iterates at the next level. Fish that are measured for length, but not age, are then assigned an age using the ALK with the highest level of specificity. HUC = Hydrologic Unit Code.

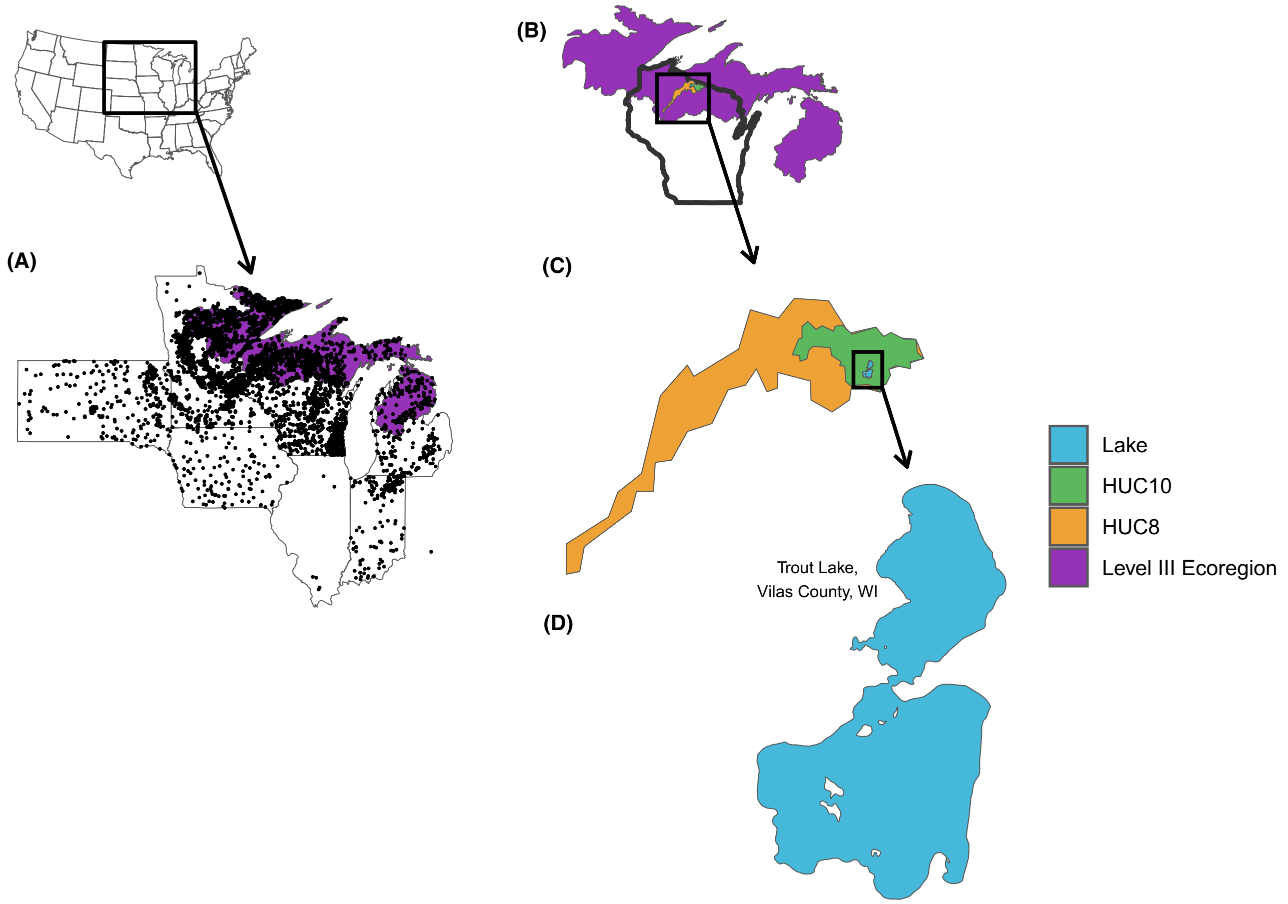

Figure 2. Lakes (A) and spatial levels (B,C) used in simulations to test the hierarchical age–length key method. An example of a single lake (D;Trout Lake, Vilas County, Wisconsin) and its associated hierarchical levels are color-coded to the levels presented in Figure 1. Age–length keys were created at the different spatial levels shown here by aggregating all paired age-length data from within a given level either across time (lake level) or across space (HUC [Hydrologic Unit Code] watershed and ecoregion levels).

survey data across 7 states, 5,279 lakes, 1,080 Hydrologic Unit Code (HUC) 10 watersheds, 251 HUC8 watersheds, and 19 Level III ecoregions (Figure 2). All data in this study came specifically from lakes, so we use that specific term in place of the more general "water body" or "population" when describing methods and results. We used paired age–length data for seven species:

- Black Crappie *Pomoxis nigromaculatus*,
- Bluegill *Lepomis macrochirus*,
- Largemouth Bass *Micropterus nigricans*,
- Northern Pike *Esox lucius*,
- Smallmouth Bass *M. dolomieu*,
- Walleye *Sander vitreus*, and
- Yellow Perch *Perca flavescens*.

In order to be included in a HALK, fish records required age, length, a water body identification code, and location (i.e., latitude–longitude, for defining watershed and ecoregion levels). We did not include sex of fish as part of these criteria, as many fish in the data set we used were either unsexed or sex was unknown. The fish from which these paired age–length data were obtained were captured using a variety of different sampling gears employed at different times of the year. Selectivity of gear type can bias sampled age–length data (Frater and Stefansson 2019), but for the purposes of presenting the HALK method, we chose to use data from all gear types in order to have a sufficient number of populations on which to assess the method. Additionally, combining samples from multiple gear types can alleviate some of these issues (Wilson et al. 2015). Several different calcified structures were used for observed age assignment, and since aging structures each have their own distinct bias, we created separate HALKs for scales, spines, and otoliths. Additionally, using age readings from multiple readers across multiple agencies could introduce bias as well; however, for the purposes of introducing the HALK concept, we assume that bias in reading age structures will balance out across so many different readers. If a HALK is to be used in practice, care is advised to ensure that bias in reading age structures is taken into consideration in HALK creation. For initial HALK creation, we used the following levels: "Lake–year" (i.e., lake–year ALK—data from a single lake in a single year), "Lake" (i.e., all years of data from a single lake), "HUC10 Watershed," "HUC8 Watershed," "Level III Ecoregion," and lastly, an ALK created for the entire species using all age–length data in our data set that was available for a species. Hydrologic Unit Code watersheds are areas of water drainage that are delineated by the U.S. Geological Survey (USGS 2023). These watersheds are hierarchically nested within one another making them well-suited for use in the HALK.

Assessing the Accuracy of Parameters from the HALK Method

We performed age assignment simulations using the *halk* R package to determine how the relative accuracy of metrics for dynamic rate functions of growth, mortality, and recruitment changed as ages were borrowed from increasingly broader strata. These simulations were conducted for each species using only lake–years where sufficient age data were collected to create lake–year ALKs. We defined data sufficiency for ALK creation as having at least five age groups and five samples per age. Defining minimum sample sizes by age is different than the commonly used minimum number of samples per length bin (Bonar et al. 2017); however, we used age for sufficiency criteria to ensure that enough age groups were present to adequately create ALKs across multiple ages at each level. Length-stratified subsamples are commonly collected in the field because it is impossible to know the age of the fish while sampling; however, the HALK uses data that are already sampled and aged. Since the point of the HALK is to share data—potentially across a wide region—we wanted to ensure that enough age groups were present to adequately assign age to fish that were measured in other water bodies. Simulations revealed that error in age assignment decreased rapidly with number of age groups used until about five, after which error only decreased marginally (Box 1 and Supplementary Materials).

As a baseline for relative accuracy, we compared growth, mortality, and recruitment metrics from HALK levels to those calculated from ages assigned by lake–year specific ALKs. Using only the observed ages from a subsample of fish would not provide an appropriate baseline for relative accuracy as these data are length-stratified and not representative of the actual length (and age) distribution for the overall sample. Previous research has shown that using only observed age data from fish in the length-stratified subsample can bias growth curve estimates (Bettoli and Miranda 2001; Goodyear 2019). Since lake–year ALKs served as the reference to compare growth and mortality metrics against, we wanted to determine accuracy and precision of these metrics at the lake–year level. To do this we performed a simulation using the estimated growth and mortality parameters from lake–year ALKs to simulate age–length-structured populations. We sampled these simulated populations for length and then subsampled them for age, doing this once for each lake–year with sufficient age data to create a lake–year ALK. We assigned age to the length sample using an ALK created from the subsample and estimated growth and mortality metrics from these assigned ages. These simulations are described in greater detail in the Supplementary Materials, and estimated growth and mortality values from them are displayed as “Lake–year (Simulated)” in the results section.

Growth was described using mean length-at-age and by fitting a Schnute parameterization of the von Bertalanffy growth curve (Schnute and Fournier 1980) using maximum likelihood in the R package TMB (Thygesen et al. 2017). The Schnute growth curve is represented by the following equation:

$$L_a = L_1 + \left[\left(L_2 - L_1 \right) \left(\frac{1 - e^{-k(a - t_1)}}{1 - e^{-k(t_2 - t_1)}} \right) \right] \tag{3}$$

Box One of the greatest concerns when using a method like the hierarchical age–length key (HALK) is error in age assignment across levels. Our results show that the age assignment error does increase as the level of data borrowing becomes less specific. However, the number of age groups used to create an ALK can also have an impact on error. More age groups will typically be included as more data are borrowed, and this trade-off captures the essence of issues that the HALK method was designed to surmount. We performed a bootstrap simulation to illustrate how age assignment error changes across both HALK level and number of age groups. Root mean squared error (RMSE) of assigned ages compared to observed ages was used to assess error in these simulations. The RMSE was highest when only a few age groups were used to create an ALK, and also increased as HALK levels increased in breadth. More age groups will generally be included as scale of data borrowing increases. If a particular lake–year only has three age groups with which to create an ALK it may be more desirable to borrow data for ALK creation at the level of Lake or Hydrologic Unit Code (HUC) 10 if six or seven age groups can be included by doing so.

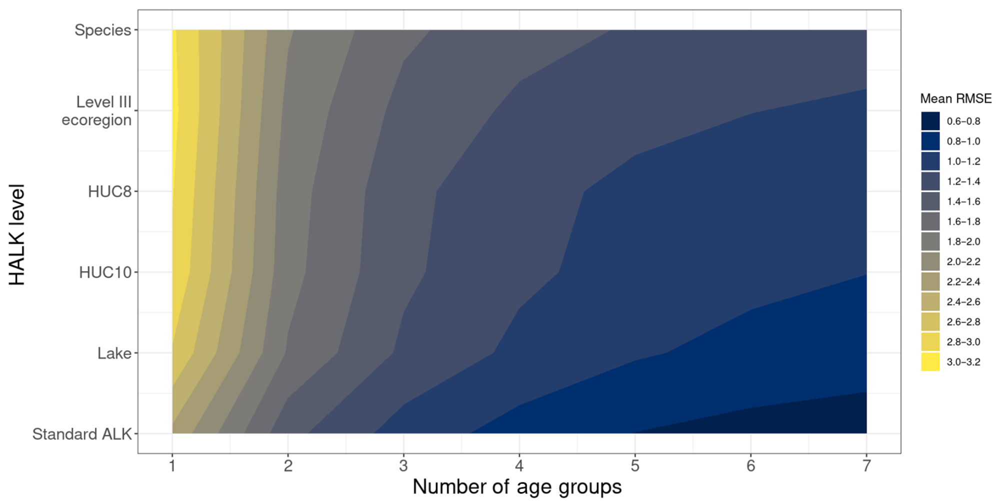

Figure B1. Contour Plot of RMSE across bootstrap simulations using a different number of age groups to create ALKs at each level of the HALK used in this model. RMSE decreases as more age groups are used and at more specific level of data-borrowing, but more age groups can typically be attained for use in an ALK by borrowing data across levels.

where L_1 is the length at minimum observed age t_1, L_2 is the length at maximum observed age t_2, and k is the usual Brody growth coefficient of the von Bertalanffy growth curve. We used a Schnute curve as parameters are uncorrelated and estimation of the function was more stable than that of the von Bertalanffy growth curve (95.5% convergence compared to 90.7%). Total instantaneous mortality (Z) was calculated using a catch-curve analysis by fitting a linear regression through the log number of fish at each age against age. Only ages greater than the most abundant age in each sample were used in catch-curve analysis. Z was converted to total annual mortality (A) using the formula $A = 1 - e^{-Z}$. Recruitment was assessed using the recruitment variability index as described by Isermann et al. (2002). We report results for both growth and mortality since these are heavily used in a management context but refer inquisitive readers to the Supplementary Materials for recruitment, where we report recruitment variability index.

We used a series of leave-one-out cross validations to estimate growth and mortality metrics when ages were assigned using ALKs constructed from increasingly broader levels of time and space. Estimates from each of these levels were compared to those of the lake–year ALK. At every HALK level, observed age data for a specific lake–year were "left out" of ALK creation and borrowing was subsequently used to assign ages to the entire sample of measured fish for that specific lake–year. The first level of the HALK (i.e., within a lake) assigned ages by pooling age–length data from all years available for a single lake minus the specific year being left out (i.e., a leave-one-year-out cross validation). At all other subsequent HALK levels, data from the particular lake of interest was left out and an ALK was created using data from all other lakes within that particular level (i.e., a leave-one-lake-out cross validation). At each level, growth and mortality metrics were estimated as described above and compared against estimates calculated using the lake–year ALK (Box 2). There were a number of cases where age data were insufficient in other years or water bodies for a given HALK level, and these instances were simply not included in that specific comparison (i.e., no level-specific ALK was created). To quantify how closely the entire growth curves from each HALK level matched those of the lake–year ALK, we used a novel approach that we refer to as the integral quotient method. This method computes percent error by dividing the area between the HALK and lake–year ALK growth curves by the area under the lake–year ALK curve. A smaller percent error represents curves that more closely resemble each other (see Appendix S1 for details).

Box To illustrate a case study from our data set we chose a randomly selected lake and year, and calculated growth, mortality, and mean length-at-age. Leigh Flowage in northeastern Wisconsin was sampled for Largemouth Bass in 2006 using a combination of mini-fyke nets and boom shockers. A total of 226 fish were sampled, and a length-stratified sub-sample of 101 age structures were collected (both scales and dorsal fin spines). The figures below show the respective metrics calculated for each level in the hierarchical age–length key (HALK) used this paper.

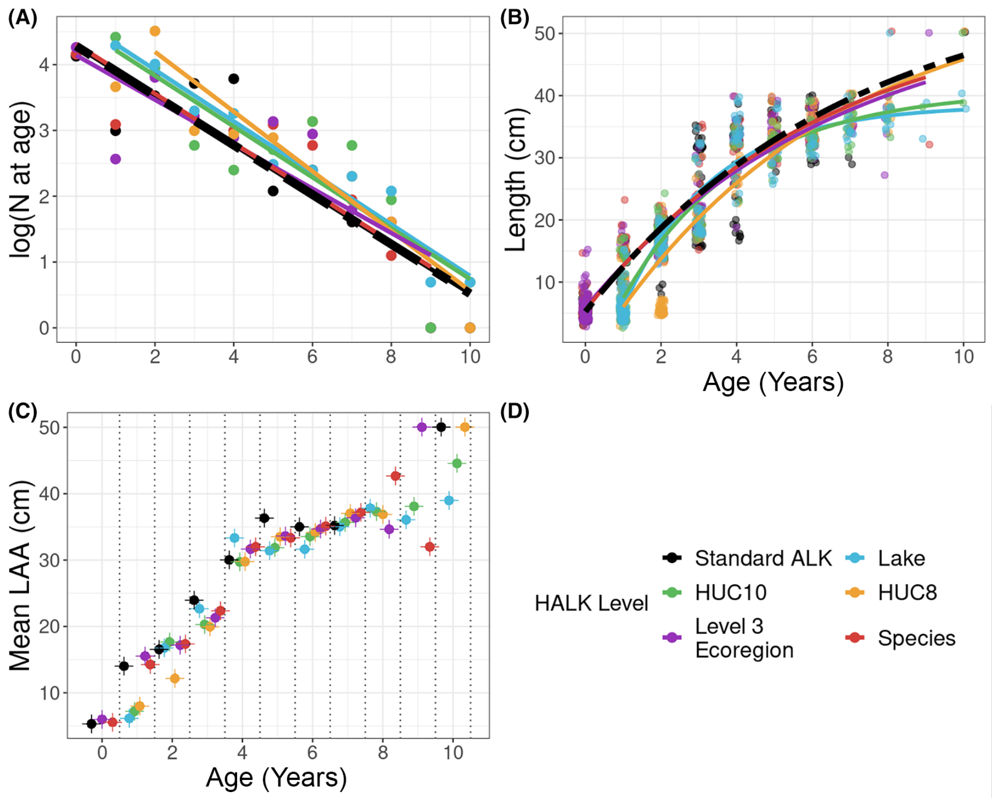

Figure B2. Mortality (**A**), growth (**B**), and mean length-at-age (**C**) for Largemouth Bass from a randomly selected lake and year (Leigh Flowage, Oconto County, Wisconsin, 2006). Values (and curves) are presented using a standard age–length key (ALK) as well as estimated ages from all different levels of the HALK. HUC = Hydrologic Unit Code.

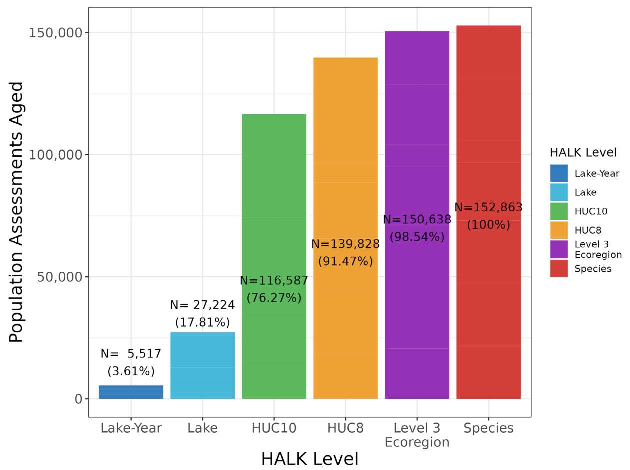

Figure 3. Number and percentage of lake–year and species-specific population assessments that could be aged at each hierarchical age–length key (HALK) level in our data set. The *N* in each bar is the total number of population assessments that could have ages assigned at each HALK level. The percentage is that number divided by the total number of population assessments.

RESULTS

A total of 5,517 unique species–lake–year combinations existed for which ages could be assigned using a lake–year ALK. Bluegill had the greatest number of lake–year ALKs at 1,561 and Smallmouth Bass had the fewest at 242. Figure 3 shows the number of species–lake–year combinations in our data set where ages could be assigned at each HALK level. By borrowing data across time within a lake the number of lake–year specific surveys that could be aged increased to nearly five times as many as could be aged using a lake–year ALK. At the HUC10 level that number rose to 116,587—over 76% of all lake–year specific surveys—and at the HUC8 level over 90% of species-specific lake–year surveys could have assigned ages using the HALK.

Cross validation revealed that the accuracy of life history metrics from ages assigned using borrowed data were generally highest at the most specific HALK level and typically lost accuracy and precision as spatial scale of borrowing increased. Some species and age-specific differences existed among different aging structures, and these are shown in the Supplementary Materials. All results presented here are from HALKs using separate aging structures, but deviations and errors are combined for simplicity in the figures. Median

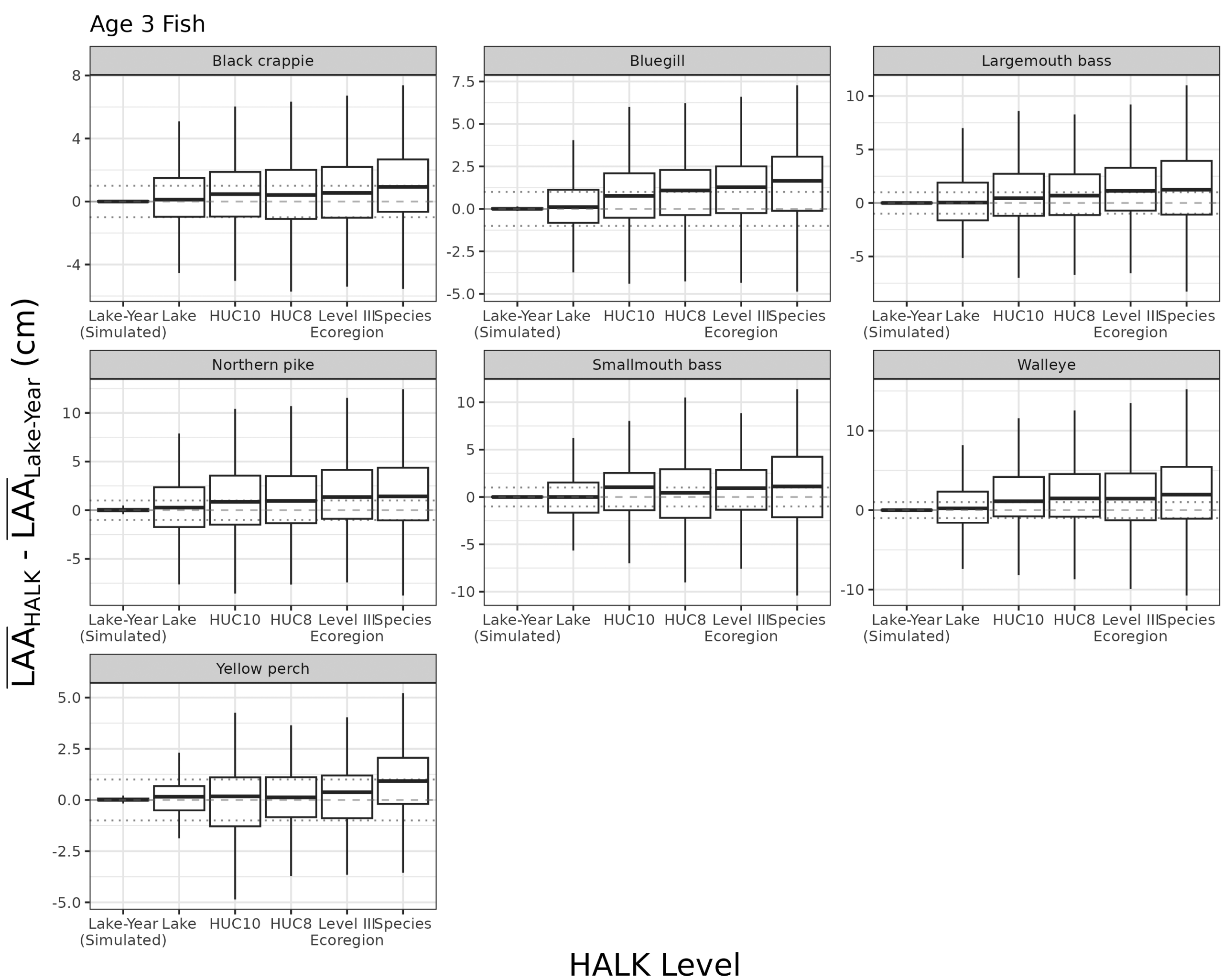

Figure 4. Deviation in mean LAA (length at age) 3 between ages assigned from hierarchical age–length key (HALK) levels and those of the lake-year ALK. The gray dotted lines indicate a deviation of 1 cm from the lake–year ALK estimate. Note that the scale of *y*-axes differ among species.

deviation in mean length of age-3 fish was 1 cm for all species at the lake level and most species at the watershed levels (Figure 4). Yellow Perch maintained median deviation in mean length at age-3 close to 1 cm for all HALK levels, except the species level as well as a low variance in the overall distribution of these values (interquartile range mostly within or near 1 cm). Median deviation in mean length of age-3 fish drifted outside of the 1 cm mark for almost all other species at HALK ecoregion level as well as the species level, and the overall distribution of these values also showed the most variance at these levels.

Deviation in total annual mortality (A) showed trends similar to those of mean length atage. Certain species, such as Northern Pike, Walleye, and Largemouth Bass, maintained median deviations of A near 0 across all HALK levels, and variance in these distributions remained relatively constant as well (Figure 5). Black Crappie, Smallmouth Bass, and Yellow Perch all had median A deviations that shifted away from 0 at certain levels, but without a consistent pattern in bias across level. However, variance in median A deviations for these species generally increased as spatial scale of data borrowing increased. Bluegill exhibited a median deviation in A close to 0 at the lake level, but began to shift from 0 at the HUC10 level and continued to be negatively biased at subsequent levels.

Percent error in growth curves exhibited greatest accuracy at the lake level, and error increased at HALK levels beyond this (Figure 6). The median percent error of growth curves was within 10% of lake–year ALK growth curves for all species at the lake level, and for most species at the watershed levels. Only Bluegill had median percent error in growth curves of less than 10% at HUC10 and HUC8 watershed levels. Northern Pike maintained consistently low percent error in growth curves across all HALK levels. Median percent error for growth curves of Northern Pike was 5.2% at the lake level and maintained consistently low error out to the species level, which was 6.2%—lower than the percent error for most HALK levels in any centrarchid. Largemouth Bass showed median percent error in the growth curve that was consistently 7.2% or less across HALK levels of lake, HUC10, and HUC8, and only increased to 8.3% and 8.6% error at the Level III Ecoregion and species levels, respectively. Yellow Perch, in contrast, showed the lowest median percent error in the growth curve for the Lake level at 3.8%, but that error more than doubled to 9.3% at the HUC10 level and then actually decreased to 8.4% at the HUC8 level. This same trend occurred for Smallmouth Bass as well, which had a median percent error of 6.4% at the lake level that increased to 9.6%

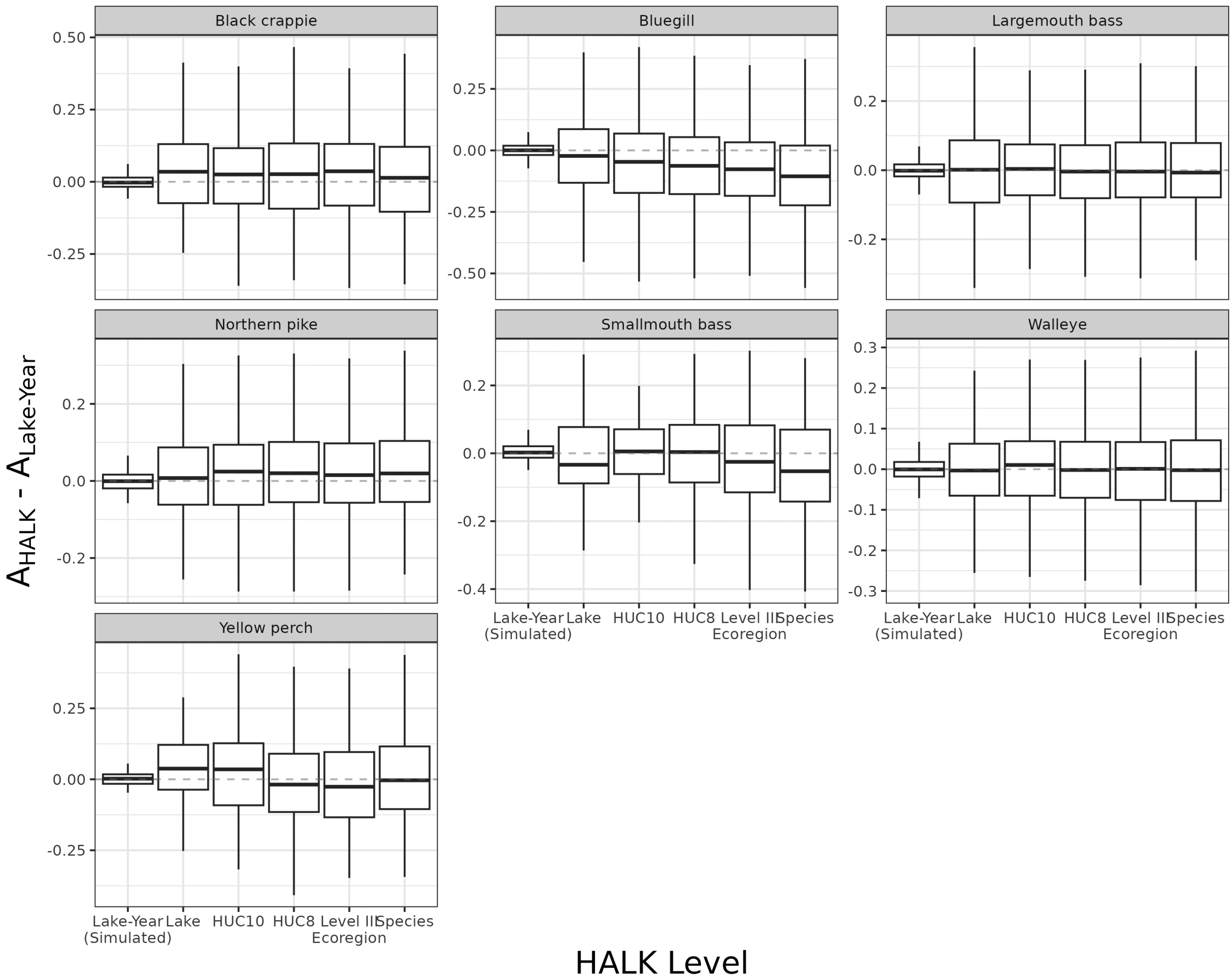

Figure 5. Deviation of total annual mortality derived from a catch-curve analysis between each hierarchical age–length key (HALK) level and the lake–year ALK. Values closer to 0 represent annual mortality estimates closer to that of the lake–year ALK. Note that the scale of *y*-axes differ among species.

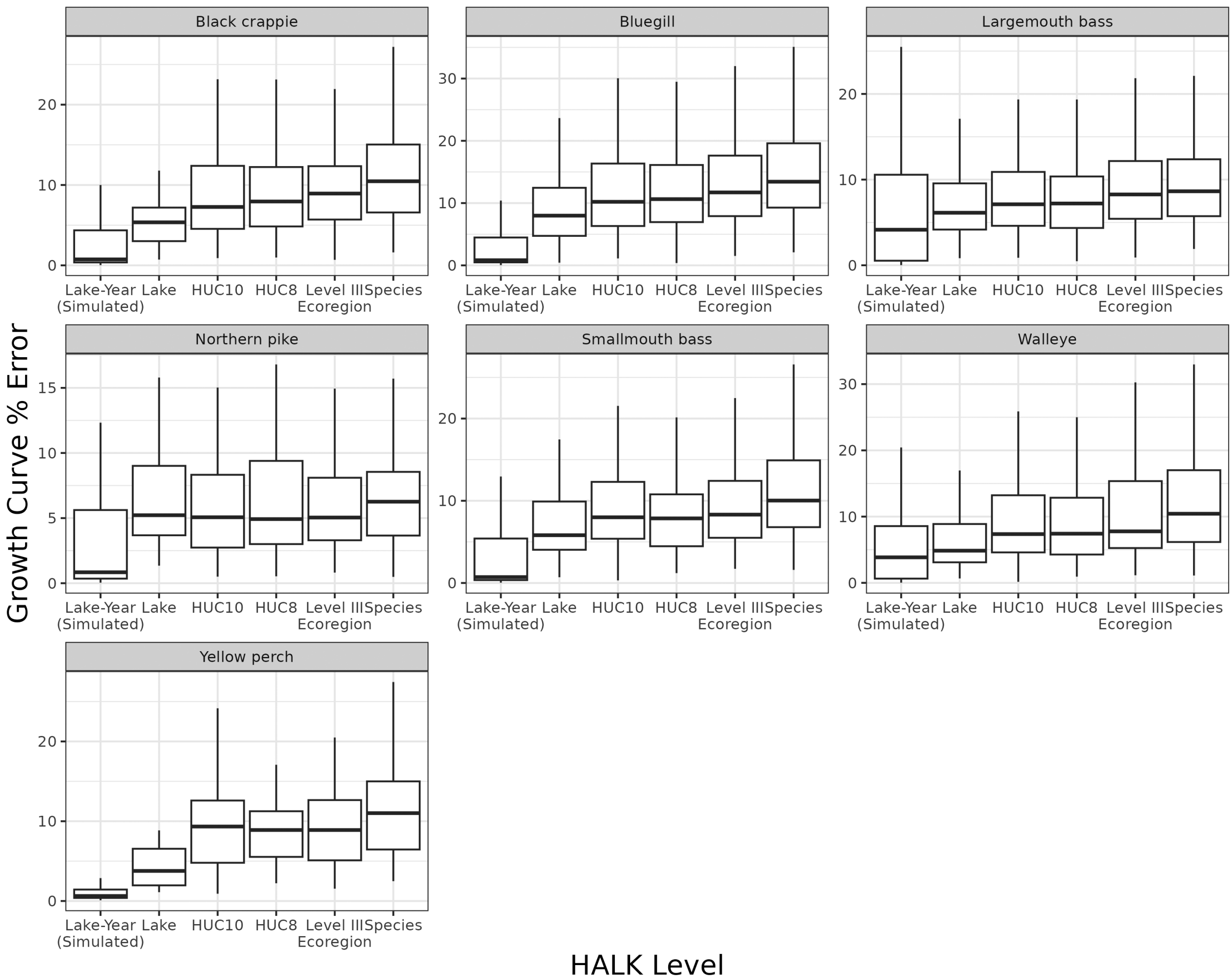

Figure 6. Percent error in growth curve between hierarchical age–length key (HALK) level and the lake-year ALK. Box and whisker plots show the distribution of errors for each species across each level. Note that the scale of *y*-axes differ among species.

at the HUC10 level and decreased to 8.9% for both the HUC8 and ecoregion levels. Walleye followed the most consistent trend of accuracy loss in growth curves across spatial scale of data borrowing. Median percent error of the growth curve for Walleye was 4.9% at the lake level, which steadily increased across each level until reaching 10.4% at the species level.

DISCUSSION

Borrowing existing age data using the HALK approach may provide opportunities for calculating age-based population metrics when lake–year-specific age data are not available—an issue that can occur for many population assessments. For example, almost 80% of lake–year-specific population surveys in our data set lacked any age data whatsoever, and only 3.6% of all population surveys had age data sufficient to create a lake–year ALK based on our criteria. These data still harbor useful information, but samples cannot be aged retrospectively. Agencies that manage water bodies are often asked to do more with less, and the HALK method and associated software that we present here aims to meet that demand by using data that are already available to gain insight into age structure of managed fish populations. To reiterate a point that we wish to make explicitly clear, this method is not a replacement for well-subsampled age data and a lake–year specific ALK; however, out of the 152,863 lake–year population surveys used in our study, only 6% had sampled for age data to create a lake–year specific ALK. This method allows scientists and managers to gain insight into age structure for the remaining 94% of lake–year surveys.

The HALK represents a method for gaining age information on these past surveys. Additionally, the HALK leverages fisheries management resources by promoting the use of age data beyond the lake and year in which collection was done. If data can be recycled, resources can perhaps be allocated elsewhere for greater impact. Properly validated data borrowing methods are increasingly widely used in fisheries science and may represent an efficient use of the best scientific information available, rather than an unfortunate stopgap measure (Hilborn and Liermann 1998; Thorson et al. 2015).

Accuracy and precision of estimated growth and mortality parameters using HALK-assigned ages declined as the spatial scale of data borrowing increased; however, the absolute values in these trends varied among species. Northern Pike and Walleye performed relatively well in terms of growth and mortality metrics across most levels of data borrowing, whereas Bluegill and Smallmouth Bass had relatively lower accuracy and less precision in growth and mortality estimates. The data set used in this study was heavily dominated by lakes in the northern regions of Wisconsin and Minnesota. This is a lake-rich area where coolwater species like Walleye and Northern Pike are relatively more common than further south in our study area, which could potentially have an influence on age assignment using borrowed data for ALK creation. For example, Walleye in a lake in northern Wisconsin are more likely to have a nearby lake where age data were collected on other Walleye, whereas Bluegill from central Illinois may not have any other data nearby that could be borrowed resulting in age assignment using data borrowed from a less-specific region like ecoregion or species. Additionally, many Bluegill samples from farther north in the species' range—where growth and mortality rates are likely different due to environmental conditions—could have a relatively large impact on age assignment when data are borrowed at greater spatial scales. This potential explanation for some of the differences among species makes the assumption that rates of either growth or mortality may be spatial autocorrelated with water bodies nearer to each other exhibiting more similar rates in either growth or mortality; however, other mechanisms may be driving differences exhibited among species. For example, Bluegill typically make up a harvest-oriented fishery (Feiner et al. 2020) and are a species that suffer from density-dependent growth (i.e., stunting; Spotte 2007). Both of these factors could play a role in altering length at age in nearby lakes, which may help explain why Bluegill exhibited relatively lower accuracy and precision in growth and mortality metrics compared to other species.

The maximum appropriate level at which to borrow data for age assignment will depend on the species, the question at hand, the tolerance for error in metrics calculated from assigned ages, and the resources and prioritization for collecting lake–year-specific age data. It is important to remember that these results represent a trade-off—having some age information of potentially lower accuracy versus having no age information at all. The HALK approach can also be contrasted with the commonly used alternative of taking a species-wide average from the literature or a global data repository. For example, the median k parameter for Walleye data in FishBase is 0.30 (Froese and Pauly 2019), which is toward the high end of k estimates for Walleye in our data set (85th percentile). The median k estimate for Walleye using lake–year ALKs in our study was 0.17.

The technique of borrowing data is not new to fisheries (Jiao et al. 2011; Punt et al. 2011; Prince et al. 2015), but most previous work has focused on borrowing from well-studied species to inform data-poor species. While this is not the first time a hierarchical approach has been used to assign ages, the previous attempt that we are aware of assigned ages to a single species (Walleye) within a portion of a single state (northern Wisconsin; Embke et al. 2019). Consequently, to our knowledge, the HALK framework presented here is the first attempt to standardize the approach and provide a framework and software for grouping data across freshwater lakes to create ALKs. Attempts have been made to implement spatial smoothing functions in ALKs for contiguous areas in marine systems (Berg and Kristensen 2012; Babyn et al. 2021). Our approach is wholly different in that it attempts to allow ages to be assigned that previously would have been impossible. The spatial ALKs of Berg and Kristensen (2012) and Babyn et al. (2021) model age using smoothing functions given length and geographical position, and they have been found to outperform traditional ALKs when applied to commercially harvested marine species. However, Aanes and Vølstad (2015) assessed ALKs from across different trawls and found that borrowing data from other gear types caused bias in the proportion-at-age. Similarly, Gerritsen et al. (2006) used a multinomial logistic model to show that ALKs vary across regions in Haddock *Melanogrammus aeglefinus* in the seas west of Scotland, suggesting that creating an ALK for all regions combined might not be the best approach. Westrheim and Ricker (1978) warned against using age–length keys from 1 year to completely inform age from another, but we found borrowing across time within a lake to have the greatest accuracy and precision in growth and mortality estimates among HALK levels. Combining data appropriately can lead to great management successes (Kolb et al. 2013), but care is advised when borrowing data from different sources, and cautionary tales depicting the ramifications of misuse abound in Katz et al. (2019). Despite many advantages, the HALK method is not a silver bullet for data-poor scenarios, and we cannot guarantee that it will work in any given situation. We suggest users assess HALK age assignment on their own data, if possible, to get a sense for how borrowing might impact age assignment error within their system.

In this study, we have presented the concept of the HALK and introduced the *halk* R package, which provides an intuitive and flexible framework for users to be able to create HALKs of their own. However, we suggest that any HALK be thoroughly vetted and tested prior to use. We have shown here that accuracy and precision increase with specificity in the level of data borrowing, but our simulations are simply to illustrate the concept. More thorough testing needs to be performed on the method of the HALK before it is more widely used. Fisheries scientists that desire to create HALKs of their own are advised to take measures to assess those HALKs using their own data before they are to be used for age assignment.

The HALK represents a novel data borrowing technique that assigns ages to fish populations to create data-rich scenarios. With further testing, this method could open the door to new possibilities with respect to optimal sampling and use of collected age data as well as the possibility to assess historical trends and patterns on existing fisheries surveys for analyses that require age data.

ACKNOWLEDGMENTS

This project was made possible by funding from the USGS National Climate Adaptation Science Center Project G21AC10338, the USGS Midwest Climate Adaptation Science Center Projects G20AC00457 and G20AC00096, the U.S. Fish and Wildlife Federal Aid in Sportfish Restoration, and the Wisconsin Department of Natural Resources. We would like to especially acknowledge the state agencies and all staff that collected and provided data used in this project: Minnesota Department of Natural Resources, Wisconsin Department of Natural Resources, Michigan Department of Natural Resources, South Dakota Game, Fish, and Parks, Iowa Department of Natural Resources, Illinois Department of Natural Resources, and Indiana Department of Natural Resources. Thank you to Mike Verhoven, Holly Kundel, and Jenna Ruzich for requesting and organizing all state agency data. These data are available on USGS ScienceBase (https://bit.ly/3R0OzHK). We would also like to sincerely thank Holly Embke for a helpful review of the paper as well as two anonymous reviewers. In addition, we wish to thank Holly Embke, Zachery Driscoll, and Derek Ogle for early conversation that led to the implementation of this method. There is no conflict of interest declared in this article.

ORCID

Paul N. Frater https://orcid.org/0000-0002-7237-6563
Zachary S. Feiner https://orcid.org/0000-0001-7880-0778
Gretchen J.A. Hansen https://orcid.org/0000-0003-0241-7048
Daniel A. Isermann https://orcid.org/0000-0003-1151-9097
Alexander W. Latzka https://orcid.org/0000-0002-3969-5714
Olaf P. Jensen https://orcid.org/0000-0001-7850-6616

REFERENCES

Aanes, S., and J. H. Vølstad. 2015. Efficient statistical estimators and sampling strategies for estimating the age composition of fish. Canadian Journal of Fisheries and Aquatic Sciences 72:938–953.

Ailloud, L. E., and J. M. Hoenig. 2019. A general theory of age–length keys: combining the forward and inverse keys to estimate age composition from incomplete data. ICES Journal of Marine Science 76:1515–1523.

Babyn, J., D. Varkey, P. Regular, D. Ings, and J. M. Flemming. 2021. A gaussian field approach to generating spatial age length keys. Fisheries Research [online serial] 240:105956.

Berg, C. W., and K. Kristensen. 2012. Spatial age–length key modelling using continuation ratio logits. Fisheries Research 129:119–126.

Bettoli, P. W., and L. E. Miranda. 2001. Cautionary note about estimating mean length at age with subsampled data. North American Journal of Fisheries Management 21:425–428.

Bonar, S. A., N. Mercado-Silva, W. A. Hubert, T. D. Beard Jr., G. Dave, J. Kubečka, B. D. Graeb, N. P. Lester, M. Porath, and I. J. Winfield. 2017. Standard methods for sampling freshwater fishes: opportunities for international collaboration. Fisheries 42:150–156.

Embke, H. S., A. L. Rypel, S. R. Carpenter, G. G. Sass, D. Ogle, T. Cichosz, J. Hennessy, T. E. Essington, and M. J. Vander Zanden. 2019. Production dynamics reveal hidden overharvest of inland recreational fisheries. Proceedings of the National Academy of Sciences of the United States 116:24676–24681.

Feiner, Z. S., M. H. Wolter, and A. W. Latzka. 2020. "I will look for you, I will find you, and I will [harvest] you": persistent hyperstability in Wisconsin's recreational fishery. Fisheries Research [online serial] 230:105679.

Francis, R. C. 2016. Growth in age-structured stock assessment models. Fisheries Research 180:77–86.

Frater, P. 2023. HALK age assignment for Midwestern lakes. Available: https://bit.ly/3RuauJ6. (November 2023).

Frater, P. N. 2020. Issues and importance of age and length data in fisheries models. Doctoral thesis. University of Iceland, Reykjavík.

Frater, P. N., and G. Stefansson. 2019. Aspects of both growth and selectivity affect growth parameter estimation bias. Fisheries Research 212:154–161.

Fridriksson, Á. 1934. On the calculation of age-distribution within a stock of cod by means of relatively few age-determinations as a key to measurements on a large scale. ICES Marine Science Symposia 86:1–14.

Froese, R. 2004. Keep it simple: Three indicators to deal with overfishing. Fish and Fisheries 5:86–91.

Froese, R., and D. Pauly, editors. 2019. FishBase [online database]. Available: www.fishbase.org. (December 2019).

Gerritsen, H. D., D. McGrath, and C. Lordan. 2006. A simple method for comparing age–length keys reveals significant regional differences within a single stock of Haddock (*Melanogrammus aeglefinus*). ICES Journal of Marine Science 63:1096–1100.

Goodyear, C. P. 2019. Modeling growth: Consequences from selecting samples by size. Transactions of the American Fisheries Society 148:528–551.

Hansen, G. J., and M. L. Jones. 2008. The value of information in fishery management. Fisheries 33:340–348.

Hilborn, R., and M. Liermann. 1998. Standing on the shoulders of giants: learning from experience in fisheries. Reviews in Fish Biology and Fisheries 8:273–283.

Hilborn, R., and C. J. Walters. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Springer Science and Business Media, Dordrecht, the Netherlands.

Isermann, D. A., and C. T. Knight. 2005. A computer program for age–length keys incorporating age assignment to individual fish. North American Journal of Fisheries Management 25:1153–1160.

Isermann, D. A., W. L. McKibbin, and D. W. Willis. 2002. An analysis of methods for quantifying crappie recruitment variability. North American Journal of Fisheries Management 22:1124–1135.

Jiao, Y., E. Cortés, K. Andrews, and F. Guo. 2011. Poor-data and data-poor species stock assessment using a Bayesian hierarchical approach. Ecological Applications 21:2691–2708.

Katz, S. L., K. A. Barnas, M. Diaz, and S. E. Hampton. 2019. Data system design alters meaning in ecological data: salmon habitat restoration across the U.S. Pacific Northwest. Ecosphere [online serial] 10:e02920.

Kerns, J., and L. Lombardi-Carlson. 2017. History and importance of age and growth information. Pages 1–8 *in* M. C. Quist and D. A. Isermann, editors. Age and growth of fishes: principles and techniques. American Fisheries Society, Bethesda, Maryland.

Kolb, T. L., E. A. Blukacz-Richards, A. M. Muir, R. M. Claramunt, M. A. Koops, W. W. Taylor, T. M. Sutton, M. T. Arts, and E. Bissel. 2013. How to manage data to enhance their potential for synthesis, preservation, sharing, and reuse—a Great Lakes case study. Fisheries 38:52–64.

Lee, H.-H., M. N. Maunder, K. R. Piner, and R. D. Methot. 2011. Estimating natural mortality within a fisheries stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. Fisheries Research 109:89–94.

Lorenzen, K., I. G. Cowx, R. Entsua-Mensah, N. P. Lester, J. Koehn, R. Randall, N. So, S. A. Bonar, D. B. Bunnell, P. Venturelli, S. D. Bower, and S. J. Cooke. 2016. Stock assessment in inland fisheries: a foundation for sustainable use and conservation. Reviews in Fish Biology and Fisheries 26:405–440.

Maceina, M. J., J. Boxrucker, D. L. Buckmeier, R. S. Gangl, D. O. Lucchesi, D. A. Isermann, J. R. Jackson, and P. J. Martinez. 2007. Current status and review of freshwater fish aging procedures used by state and provincial fisheries agencies with recommendations for future directions. Fisheries 32:329–340.

Magnusson, A., and R. Hilborn. 2007. What makes fisheries data informative? Fish and Fisheries 8:337–358.

Mildenberger, T. K., M. H. Taylor, and M. Wolff. 2017. TropFishR: an R package for fisheries analysis with length-frequency data. Methods in Ecology and Evolution 8:1520–1527.

MNDNR (Minnesota Department of Natural Resources). 2017. Manual of instructions for lake survey. MNDNR, Special Publication 180, St. Paul. Available: https://bit.ly/49mG2aM. (November 2023).

Ogle, D. H., J. C. Doll, A. P. Wheeler, and A. Dinno. 2023. FSA: simple fisheries stock assessment methods. Available: https://bit.ly/40ndMjY. (November 2023).

Oke, K., C. Cunningham, P. Westley, M. Baskett, S. Carlson, J. Clark, A. Hendry, V. Karatayev, N. Kendall, J. Kibele, H. K. Kindsvater, K. M. Kobayashi, B. Lewis, S. Munch, J. D. Reynolds, G. K. Vick, and E. P. Palkovacs. 2020. Recent declines in salmon body size impact ecosystems and fisheries. Nature Communications [online serial] 11:4155.

Ono, K., R. Licandeo, M. L. Muradian, C. J. Cunningham, S. C. Anderson, F. Hurtado-Ferro, K. F. Johnson, C. R. McGilliard, C. C. Monnahan, C. S. Szuwalski, J. L. Valero, K. A. Vert-Pre, A. R. Whitten, and A. E. Punt. 2015. The importance of length and age composition data in statistical age-structured models for marine species. ICES Journal of Marine Science 72:31–43.

Prince, J., A. Hordyk, S. R. Valencia, N. Loneragan, and K. Sainsbury. 2015. Revisiting the concept of Beverton–Holt life-history invariants with the aim of informing data-poor fisheries assessment. ICES Journal of Marine Science 72:194–203.

Punt, A. E., D. C. Smith, and A. D. Smith. 2011. Among-stock comparisons for improving stock assessments of data-poor stocks: the "Robin Hood" approach. ICES Journal of Marine Science 68:972–981.

Quinn, T. J., and R. B. Deriso. 1999. Quantitative fish dynamics. Oxford University Press, Oxford, UK.

Quist, M. C. 2007. An evaluation of techniques used to index recruitment variation and year-class strength. North American Journal of Fisheries Management 27:30–42.

Schnute, J., and D. Fournier. 1980. A new approach to length–frequency analysis: growth structure. Canadian Journal of Fisheries and Aquatic Sciences 37:1337–1351.

Sippel, T., H. H. Lee, K. Piner, and S. L. Teo. 2017. Searching for *M*: is there more information about natural mortality in stock assessments than we realize? Fisheries Research 192:135–140.

Solokas, M. A., Z. S. Feiner, R. Al-Chokachy, P. Budy, J. T. DeWeber, J. Sarvala, G. G. Sass, S. A. Tolentino, T. E. Walsworth, and O. P. Jensen. 2023. Shrinking body size and climate warming: many freshwater salmonids do not follow the rule. Global Change Biology 29:2478–2492.

Spotte, S. 2007. Bluegills: biology and behavior. American Fisheries Society, Bethesda, Maryland.

Thorson, J. T., J. M. Cope, K. M. Kleisner, J. F. Samhouri, A. O. Shelton, and E. J. Ward. 2015. Giants' shoulders 15 years later: lessons, challenges and guidelines in fisheries meta-analysis. Fish and Fisheries 16:342–361.

Thygesen, U. H., C. M. Albertsen, C. W. Berg, K. Kristensen, and A. Nielsen. 2017. Validation of ecological state space models using the Laplace approximation. Environmental and Ecological Statistics 24:317–339.

USGS (U.S. Geological Survey). 2023. USGS water resources: about water resources. Available: https://bit.ly/3sfCF4H. (November 2023).

WDNR (Wisconsin Department of Natural Resources). 2013. Fisheries management handbook: policies, procedures, and guidance. WDNR, Division of Fish, Wildlife, and Parks, Bureau of Fisheries Management, Madison.

Westrheim, S., and W. Ricker. 1978. Bias in using an age–length key to estimate age-frequency distributions. Journal of the Fisheries Board of Canada 35:184–189.

Wilson, K. L., B. G. Matthias, A. B. Barbour, R. N. Ahrens, T. Tuten, and M. S. Allen. 2015. Combining samples from multiple gears helps to avoid fishy growth curves. North American Journal of Fisheries Management 35:1121–1131.

SUPPORTING INFORMATION

Additional supplemental material may be found online in the Supporting Information section at the end of the article.

Appendix S1

Data S1 AFS