North American Journal of Fisheries Management 28:1259–1269, 2008

DOI: 10.1577/M07-086.1

[Article]

Evaluation of Predicted Fish Distribution Models for Rare Fish Species in South Dakota

CARI-ANN HAYER* AND STEVEN S. WALL[1]
Department of Wildlife and Fisheries Sciences, Box 2140B, South Dakota State University, Brookings, South Dakota 57007, USA

CHARLES R. BERRY, JR.
South Dakota Cooperative Fish and Wildlife Research Unit, Box 2140B, South Dakota State University, Brookings, South Dakota 57007, USA

Abstract.—Predictive fish distribution models have utility in planning conservation measures for rare fish species. However, species rarity creates sampling and modeling difficulties that require an understanding of model accuracy. We evaluated existing distribution models for 10 rare fishes based on 2,026 community fish samples and associated riverine habitat. Our fieldwork provided an independent fish species inventory for 143 sample sites. This inventory was used to quantify species detectability for use as a weighting factor with which to correct false-negative modeling errors. Presence/absence data were compared with predictions to evaluate model accuracy as determined by Cohen's kappa and correct classification rates. Detection probabilities were generally small but ranged from 0 to 0.68. The models predicted species occurrence with relatively high success (average kappa = 45.6% and average correct classification rate = 74.1%). The habitat variables for predicting species occurrence varied among species; however, stream size and streamflow were the most influential. All distribution models had adequate predictive abilities and improved our understanding of fish distributions and the factors determining those distributions. Model accuracy statistics can provide managers with a measure of confidence when they are directing conservation activities.

The conservation of rare fish species requires biogeographic information about fish distributions that is spatially comprehensive and of appropriate resolution for effective fisheries management. This need can be addressed with statistical fish distribution models based on habitat relationships that predict species occurrence (Jennings 2000; Oakes et al. 2005). Models can be useful for designing survey and monitoring programs, assessing habitat suitability, and prioritizing sites for conservation (Pearce and Ferrier 2000; Olden and Jackson 2002).

Sampling rare species can be challenging (McArdle 1990; Scott et al. 2002; MacKenzie et al. 2005), and insufficient data often preclude predictive distribution modeling (Scott et al. 2002; Rushton et al. 2004). A species may be considered rare because of low abundance, cryptic or nocturnal behavior, or temporal variation in its distribution (Gaston 1994; McDonald 2004; MacKenzie et al. 2005), all of which influence detectability. Species presence estimates are generally derived from catch data that are assumed to be representative of actual species presence in the sampled area (Schmidt 2003). However, it is unusual to detect all of the species present in a particular sampling area (Pollock et al. 2002; Gu and Swihart 2004). This is especially true for mobile aquatic animals, such as fish (Boulinier et al. 1998; Nichols et al. 1998a) in complex riverine habitats. Rare species have low model sensitivity, making it difficult to predict the occurrence of the organisms whose conservation and management may be the most critical (Olden and Jackson 2002).

Detection is defined as the probability of detecting at least one individual of a species, assuming that the species is present at the sampling site at that particular time (Boulinier et al. 1998; Nichols et al. 1998a; Bayley and Peterson 2001) and is a function of abundance, sampling efficiency, and effort (Nichols et al. 1998b; Gu and Swihart 2004). The inability to detect a species can inflate commission error (false positives [i.e., the species is predicted to occur at the site but is not actually found there]; Scott et al. 1993; Cassidy et al. 1994; Krohn 1996). Commission error has two components: true error (the species is not found because of inaccuracies in the models) and apparent error (the species does occur at the site but was missed in field surveys) (Dedon et al. 1986; Scott et al. 1993; Edwards et al. 1996). To minimize commission error, other authors have applied a 0.5 threshold as a weighting factor on commission (e.g., Karl et al. 2000), which increased the predictive ability

* Corresponding author: cari-ann.hayer@sdstate.edu
[1] Present affiliation: James River Water Development District, Box 849, Huron, South Dakota 57350, USA.

Received May 17, 2007; accepted January 21, 2008
Published online August 25, 2008

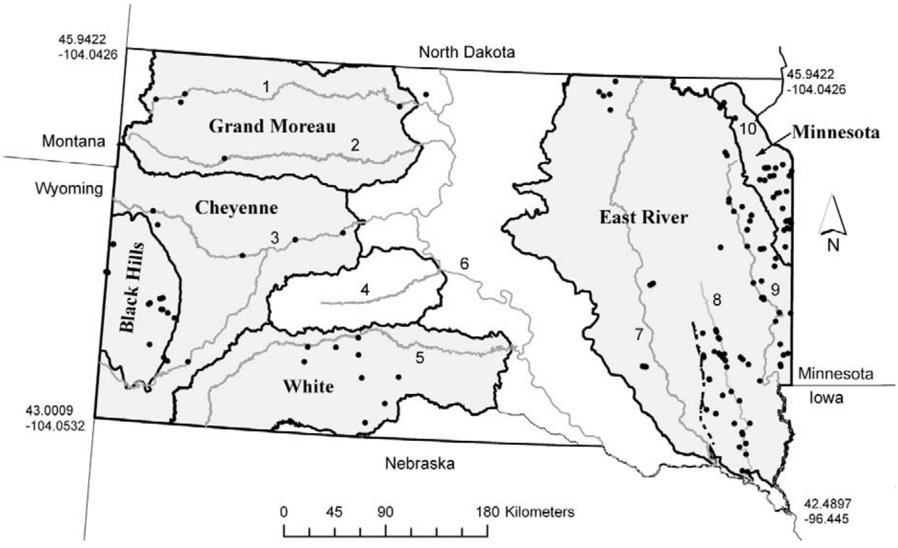

FIGURE 1.—Map of South Dakota showing the 10 major river drainages, the six ecological drainage units (EDUs) used in this study (shaded areas), and the individual data sites (black dots). The Grand–Moreau EDU consists of the Grand (1) and Moreau (2) rivers. The Cheyenne EDU includes the Cheyenne River (3) and the White EDU includes the White River (5). The East River EDU includes the James (7), Vermillion (8), and Big Sioux (9) rivers. The Minnesota EDU includes the upper Minnesota River (10). Rivers not included in this study are the Bad River (4) and the Missouri River main stem (6).

of models. However, this arbitrary number does not account for heterogeneous detection rates among species. Including detection probabilities as an error estimate (Hoeting et al. 2000) may improve model performance and minimize the bias in parameter estimates and the associated uncertainty (MacKenzie et al. 2002; Martin et al. 2005).

The utility of predictive models has been questioned because many of them lack validation and accuracy assessment (Short and Hestbeck 1995; Davis 1996; Edwards 1996; Scott et al. 1996; Guisan and Zimmerman 2000; Karl et al. 2000). Model validation involves quantifying the predictive performance of the model (e.g., determining the correct classification rate for species presence or absence). Validation is an important step in the modeling process because it quantifies our confidence in the predictions and potential model applications. To obtain unbiased estimates of the predictive accuracy of a model, data collected from sites other than those used to develop the model should be used (Pearce and Ferrier 2000; Olden et al. 2002).

Recent data showing declines in fish species in South Dakota rivers (Hoagstrom et al. 2007) and a new emphasis on conserving nongame species (SDGFP 2007) has heightened the need for tools to assist in conservation planning. Aquatic gap analysis (Sowa et al. 2007) has been suggested for this purpose because this type of analysis has provided distribution models for South Dakota fishes (Wall et al. 2004).

The goal of this study was to assess the predictive accuracy of existing gap analysis models for 10 rare fishes. We hypothesized that commission errors resulted from the failure to find a rare species when it was present and that these errors could be reduced by incorporating detection probabilities. Our objectives were to assess model accuracy across six ecological drainage units, use detection probabilities in the assessment of model accuracy (Hayer et al. 2006), and improve information about the factors that influence the distribution of these species.

Study Area

South Dakota includes 10 major river drainages (Figure 1) that lie within two geomorphic provinces, the Central Lowlands and the Great Plains (Thornbury 1965; Hogan 1995; Holliday et al. 2002). The Missouri Coteau (Missouri River main stem) forms the western edge of the Central Lowlands Province that includes the Big Sioux, Vermillion, James, and upper Minnesota rivers. The Central Lowlands Province is lower in elevation than the Great Plains Province, was entirely glaciated, and is characterized by abundant lakes, wetlands, and low-gradient streams (Hoagstrom et al. 2007).

In contrast, the Great Plains Province was glaciated only near its boundaries and is characterized by rivers and streams (Winter and Woo 1990), erosive soils, and few natural lakes (Meade et al. 1990). The landscape

TABLE 1.—Valley segment habitat variables and associated classes used in modeling the distribution of rare fish species in South Dakota.

Variable	Number of classes	Classes	Method or criterion
Stream size	4	Headwater, creek, small river, large river[a]	Shreve link
Channel slope	3	Low, medium, high	Elevation change between valley segments relative to stream size and ecological drainage unit
Network position	11	Headwater to creek, creek to large river, large river to extra large river, etc.	The connection of one stream reach to the next downstream reach, downstream Shreve link
Parent material	6	Glacial till, outwash, loess, etc.	Dominant soil parent materials
Stream flow	2	Perennial or intermittent	Classified according to the National Hydrography Database
Groundwater potential	3	Low, medium, high	Based on groundwater velocity
Connectivity to lake	2	Yes or no	Valley segments connected to water bodies >4 ha
Floodplain influence	2	Yes or no	Slope from digital elevation model to determine valley wall, valley segments >249 m within valley wall
Elevation	3	Low, medium, high	Maximum elevation of the valley segment

[a] Additional stream size categories included Strahler stream order and the nine classes obtained by including the median Shreve link values between the standard four classes. Only one stream size category was used in a model, however, namely, the one that produced the best model fit.

has greater topographic relief than the Central Lowlands Province and includes the Grand, Moreau, Cheyenne, Bad, and White river drainages in South Dakota. The Great Plains Province has a wide range of stream types as a result of variable local geology and climate (Thornbury 1965). In addition, the Black Hills is a unique landform within the Cheyenne River drainage in South Dakota that is typically characterized by clear coldwater streams with deep, narrow valleys and limited floodplains (Hogan 1995).

The major river basins in South Dakota have been partitioned into six ecological drainage units (EDUs; Figure 1). An EDU is a zoogeographic unit that has a relatively distinct physiographic character, evolutionary history, and aquatic assemblage (Higgins et al. 2005; Sowa et al. 2007). Three glaciated basins—the James, Vermillion, and Big Sioux—make up one EDU because of the similarities in their fish assemblages and habitat (Wall et al. 2004). The upper Minnesota River drainage is an EDU because it is in the Mississippi River basin, whereas other South Dakota rivers are in the Missouri River basin. The major western tributaries to the Missouri River in South Dakota have distinct zoogeographies (Hoagstrom and Berry 2006) that suggested division into a White River EDU, Cheyenne River EDU, and Grand–Moreau River EDU. The Black Hills, characterized by unique geology and coldwater streams, make up a distinct ecoregion within the northwestern Great Plains (Omernik 1987) and thus were considered an EDU for this study.

Methods

Distribution models for South Dakota fishes were available from our unpublished studies (Smith et al. 2002; Wall et al. 2004), and these models were evaluated for fishes in selected basins outside South Dakota (Sylvester 2004). The modeling process included five steps: developing the base layer for modeling, compiling a comprehensive fish database, developing an empirical model, predicting distribution, and assessing model accuracy.

Base layer for modeling.—We used geographical information systems (GIS) technology to classify riverine habitat at multiple spatial scales according to the landscape variables that are important in determining the distribution of aquatic communities (Table 1; Lammert et al. 1996; Higgins et al. 1998; Smith et al. 2002; Sowa et al. 2007) based on a hierarchy system (Frissell et al. 1986). A GIS analysis was performed with ArcView 3.2 (ESRI 1999), ArcInfo version 8.2 (ESRI 2000), and ArcInfo Arc Macro Language supplied by the Nature Conservancy (TNC–FWI 2000) and the Missouri Resource Assessment Partnership (Annis et al. 2002). The National Hydrography Dataset (USGS 2002) was used to develop the base hydrography layer (1:100,000 scale) from which valley segment types were delineated.

Fish database.—A comprehensive fish database was compiled from numerous sources (recently summarized by Hoagstrom et al. 2007). Sample points on rivers and the associated fish presence data were converted to a spatially referenced layer for use in GIS analysis. Point locations were verified for accuracy. A total of 2,026 survey sites were used for modeling. We spatially linked the data from each site to the valley segment layer. We consolidated data from sites with multiple samples at different times. Fish were considered present in a valley segment if they occurred there at any time (i.e., historic [before 1970] samples were used) to maximize the sample size for analyses. Species were verified to be within their ranges according to the fish distributions reviewed by Hoagstrom et al. (2007).

Model development.—Once fish survey sites were spatially linked to valley segments, fish presence/

TABLE 2.—Percentages of predictive models for 10 rare fishes in South Dakota in which particular habitat variables (Table 1) appear.

Species[a]	Stream size	Stream flow	Channel slope	Parent material	Groundwater potential	Connectivity to lake	Elevation	Network position
Finescale dace *Phoxinus neogaeus***	0.50	1.00	0.50	0.50				
Northern redbelly dace *Phoxinus eos**	0.67	0.80	0.15					
Central mudminnow *Umbra limi*	1.00	0.70			0.20	0.50		
Longnose sucker *Catostomus catostomus**	1.00						1.00	
Trout-perch *Percopsis omiscomaycus*	1.00							
Sturgeon chub *Macrhybopsis gelida**	1.00			0.50	0.30			
Carmine shiner *Notropis percobromus*	1.00							1.00
Pearl dace *Margariscus margarita*	1.00							
Hornyhead chub *Nocomis biguttatus*	1.00							
Banded killifish *Fundulus diaphanus***						1.00		
All species	9	3	2	2	2	2	1	1

[a] Species with one asterisk are listed as threatened in South Dakota, species with two asterisks as endangered (SDGFP 2007).

absence data were attributed to the physical features of valley segments for use in modeling (Table 1). Our modeling approach followed that of Sowa et al. (2007) and included a decision tree analysis (i.e., classification and regression tree analysis; Breiman et al. 1984). Like Sowa et al. (2007), we used AnswerTree 3.0 statistical software (SPSS 2001) to perform the decision tree analysis and applied the exhaustive chi-squared automatic interaction detector algorithm (Biggs et al. 1991), which uses chi-square statistics to identify optimal splits, for each predictor. We used the "relative 50% approach" as defined in Sowa et al. (2007) to prune our decision trees and minimize overfitting of the data, which produces trees with too many branches and increases the misclassification rate of the model (Breiman et al. 1984). The model result is a dendrogram showing a set of mutually exclusive decision rules identifying predictor variable combinations (valley segment habitat features) that are significantly associated with species occurrence. If we could not obtain an output from AnswerTree (usually because the sample size was too small), we used published valley segment affinities and habitat affinities that matched the species occurrence to predict valley segments for fish presence, as in the approach used in the South Dakota gap analysis for aquatic systems (Smith et al. 2002).

We modeled fish species separately according to three faunal regions to account for regional variations in species and habitat associations: (1) the Missouri River basin east of the Missouri River; (2) the Missouri River basin west of, and including, the Missouri River; and (3) the Minnesota River basin. The input data used to model a given species were chosen from eight-digit hydrologic unit codes (HUCs) where that species occurred to minimize the effect of distributional constraints on model performance (Sowa et al. 2007). We used different variables in each region because previous studies in other states (Wall et al. 2005; Sowa et al. 2007) have shown that the power of habitat variables to predict fish distributions varies geographically. For example, elevation and temperature were important in determining fish distributions in the Black Hills but not those in the low-gradient warmwater streams of the Central Lowlands (Hoagstrom et al. 2007). We used the regional gradient and elevation variables for analysis except when a species distribution was restricted to a major drainage, in which case we used the gradient relative to that drainage. When the original four stream size-classes (i.e., headwater, creek, small river, and large river; Table 1) were not good predictors, they were separated into smaller categories to increase variability and model accuracy. We also used Strahler order as a predictor when that produced a stronger model. To minimize collinearity among independent variables, only one attribute representing stream size was used in a model (either stream size or stream order).

Predicted distribution maps.—Predicted distribution maps were generated by entering model output for each species into a GIS query to select valley segments containing habitat features associated with occurrence of the species. Fish were predicted to be present in valley segments with habitat variables associated with the fish and were considered absent in valley segments without habitat variables associated with the fish. Model predictions were restricted to known species' ranges within the 8- or 11-digit HUC level for relict species occurring in isolated watersheds.

Model accuracy assessment.—For this report, we evaluated the distribution models of 10 rare fishes (Table 2) that have priority in conservation planning (SDGFP 2007). Distribution maps for these species were used to identify the valley segments in which that species' presence was predicted. We used an independent data set from 143 predicted sites (Figure 1) that were sampled during summer 2004 (June–September) and summer 2005 (July–September). Fish were captured with 4.7-mm-mesh seines (seine length varied from 5 to 10 m depending on stream conditions) and backpack electrofishing (Smith-Root; Model LR-24). All fish were identified to species, recorded as present, and released. These data were used to compute detection probabilities and accuracy statistics for all species (Hayer et al. 2006); however, for this paper we evaluate model accuracy only for the species of concern.

Detection probabilities (DPs) were calculated with the maximum likelihood methods described by MacKenzie et al. (2002, 2003) and recently used by others (Bailey et al. 2004; Wintle et al. 2004; Hayer and Irwin, in press). Detection probabilities were based on presence/absence data gathered in the independent data set and were calculated by EDU.

Model accuracy was evaluated at the valley segment scale for EDUs at which independent data test sites occurred (Figure 1). The models were tested separately for each EDU because of variance in species prevalence, which affects model accuracy. For example, fathead minnow *Pimephales promelas* are very abundant in most streams across South Dakota, with exception of streams at high elevations (i.e., in the Black Hills). Also, species composition varies by EDU. For example, longnose dace *Rhinichthys cataractae* occur west of the Missouri River but not east of it, where the western blacknose dace *R. obtusus* becomes the dominant *Rhinichthys* species. For species with restricted distributions, such as the northern redbelly dace, only test sites within the species' 8-digit or 11-digit distributions were used.

Fish survey sites from the independent data set were spatially linked to the valley segment layers. Valley segments were coded into four classes: true presence (the fish species was both predicted to occur in the valley segment and captured there); commission error (the species was predicted to occur in the valley segment but was not captured there); omission error (the species not predicted to occur in the valley segment but was captured there); and true absence (the species was neither predicted to occur in the valley segment nor captured there). These classes were entered into a 2 × 2 confusion matrix. The matrix was used to calculate accuracy metrics (Fielding and Bell 1997) for each fish species sampled in the independent data set.

We used Cohen's kappa (κ) as the principal indicator of model accuracy (Fielding and Bell 1997; Manel et al. 2001). Cohen's kappa is the ratio of the proportion of agreement to the maximum number of times that the predictions could agree with the observations. It measures model agreement with what is observed rather than model association with what is observed (Fielding and Bell 1997). Values of 0.0–0.4 are considered to indicate "slight to fair" model performance, values of 0.4–0.6 "moderate" performance, values of 0.6–0.8 "substantial" performance, and values of 0.8–1.0 almost "perfect" performance (Landis and Koch 1977; Fielding and Bell 1997; Manel et al. 2001). Negative values are possible, and theses indicate "poor" model agreement (Manel et al. 2001). The value of κ is thus the probability (above that of pure chance) that the model will predict the presence or absence of a species in a valley segment. In addition, we report correct classification rates (CCRs), which represent the percentage of cases that are correctly predicted.

We used detection probabilities to minimize the false-positive error rate (i.e., commission error). In the confusion matrix, commission error was recalculated by multiplying the actual number of false positives from the field surveys by the detection probability of that species. Accuracy measures were recalculated using the weighted value for commission error. Only the weighted values for accuracy measures are reported unless otherwise specified.

Results

Stream size was a key habitat variable in predicting the presence of nine species (Tables 2, 3). Certain species were associated with large rivers (e.g., sturgeon chub), whereas others were associated with headwaters and creeks (e.g., carmine shiner). Flow was important in predicting the presence of three species. Other habitat variables entered the predictive models for one or two species; the one exception was floodplain influence, which was not present in any models.

Sixty-two species were collected from 143 independent sampling sites across six EDUs (Hayer et al. 2006; Figure 1). Detection probabilities varied among species and EDUs, ranging from 0% to 100%. The DPs for the targeted rare species ranged from 0.001 (central mudminnow) to 0.68 (hornyhead chub; Table 4) and varied among EDUs. For example, the DPs for the northern redbelly dace were 0 in the Grand–Moreau and White EDUs, 0.004 in the Minnesota EDU, and 0.213 in the East River EDU. The DPs for banded killifish, finescale dace, longnose sucker, and pearl dace were all 0, as these species were not found in the independent data set.

The values of κ and CCR varied among EDUs and the 62 species. The value of κ ranged from -0.012 to $+1.00$ and that of CCR from 0 to 1.00, indicating a wide range of model accuracy; this trend was also apparent for the 10 rare species. There was moderate to

TABLE 3.—Fish species of concern in South Dakota and habitat variables used in predictive models. Ecological drainage units (EDUs) are large landscape areas with uniform biotic and abiotic features (see Figure 1).

Species	Variable(s)	EDU
Banded killifish	Stream segments connected to a lentic system or glaciated lake	Minnesota and East River
Carmine shiner	Creeks and small rivers with perennial flow and no size discrepancy or confluence of a creek with a small river	Minnesota
Central mudminnow	Creeks and headwaters with perennial flow, creeks and headwaters with intermittent flow that are connected to a lake or slough; or creeks and headwaters with intermittent flow and groundwater potential	Minnesota
	Creeks and headwaters connected to a lake or slough	East River
Finescale dace	Stream segments with intermittent flow, medium to high gradients, and stream orders exceeding 1 or stream segments with perennial flow through bedrock geology	White River and Black Hills
Hornyhead chub	Headwaters, creeks, and small rivers	Minnesota
Longnose sucker	Headwaters, creeks, and small rivers at medium elevations	Black Hills
Northern redbelly dace	Headwaters and creeks with perennial flow or with intermittent flow and medium to high groundwater potential	Minnesota
	Segments adjacent to valley segments where the fish species was present	East River
	Stream segments with intermittent flow and medium to low groundwater potential, second- to fourth-order streams with perennial flow, or fifth-order streams with perennial flow and medium to high gradients	White and Grand–Moreau
Pearl dace	Headwaters and creeks with stream orders exceeding 1	White
Sturgeon chub	Fifth-order streams flowing through alluvial outwashes or sixth-order streams	White, Cheyenne, and Grand–Moreau
Trout-perch	Fourth- and fifth-order streams	East River

substantial model agreement ($\kappa > 40\%$) for all rare species except the central mudminnow in the Minnesota EDU and the longnose sucker in the Black Hills EDU); model agreement was variable among drainages and species (Table 4). Correct classification rates for rare species were also variable, ranging from a low of 0.467 for the longnose sucker to a high of 0.955 for the banded killifish (Table 4); this range was narrower than that exhibited by all 62 species.

The model for the northern redbelly dace in the East

TABLE 4.—Accuracy metrics for species of concern in South Dakota across six ecological drainage units (EDUs). Abbreviations are as follows: DP = detection probability, CCR = correct classification rate, and κ = Cohen's kappa statistic (see text for details). Model performance is based on Landis and Koch (1977).

Species	EDU	True presences	Commission errors (false positives)	Omission errors (false negatives)	True absences	DP	CCR	κ	Model performance
Banded killifish	Minnesota	0	1	0	21	0	0.955	0.488	Moderate
	East River	0	3	0	59	0	0.952	0.488	Moderate
Carmine shiner	Minnesota	2	9	0	11	0.004	0.591	0.448	Moderate
Central mudminnow	Minnesota	0	2	1	19	0.004	0.864	0.364	Fair
	East River	1	6	0	55	0.001	0.903	0.545	Moderate
Finescale dace	Black Hills	0	3	0	12	0	0.8	0.444	Moderate
Hornyhead chub	Minnesota	11	7	0	4	0.68	0.682	0.441	Moderate
Longnose sucker	Cheyenne	0	2	0	5	0	0.714	0.417	Moderate
	Black Hills	0	8	0	7	0	0.467	0.318	Fair
Northern redbelly dace	Grand–Moreau	0	1	0	4	0	0.8	0.444	Moderate
	Minnesota	2	8	0	12	0.004	0.636	0.476	Moderate
	White	0	3	0	6	0	0.667	0.4	Fair
	East River	2	0	2	13	0.213	0.882	0.605	Substantial
Pearl dace	White	0	3	0	6	0	0.667	0.4	Moderate
Sturgeon chub	Cheyenne	1	2	0	4	0.012	0.714	0.561	Moderate
Trout-perch	East River	3	14	0	16	0.084	0.576	0.424	Moderate

River EDU had the greatest predictive accuracy (κ = 0.605, indicating substantial model performance; Table 4). The DP was 0.21 (Table 4), and habitat modeling predicted that this species would occur in river segments up- and downstream from those in which the species had previously been recorded (Table 3). We collected this species at two sites where it was predicted and two where it was not predicted. This was the only model assessment that did not produce a commission error.

The banded killifish, finescale dace, longnose sucker, and pearl dace were not detected in the independent survey; however, model accuracy was fair to moderate (Table 4), ranging from 0.318 for the longnose sucker in the Black Hills EDU to 0.488 for the banded killifish in the Minnesota and East River EDUs. Incorporating DPs of 0 into the calculation of κ had the greatest effect on these species because the number of commission errors became zero.

Incorporating DPs resulted in larger κ values for all rare fish species except the northern redbelly dace in the East River EDU, for which κ remained the same. For all 62 species, 44% of the models had κ values greater than 0.40 when DPs were included, compared with 18% of the models when they were not included; the value of κ decreased for one species when DPs were included and did not change for 15% of the 62 species.

Discussion

Our study was the first to incorporate DPs into testing model performance to reduce the bias associated with sampling fish, although others have suggested the approach (MacKenzie et al. 2002, 2005). Information on fish distributions in Alabama streams was improved by the use of DPs (Hayer 2005). The correctly predicted presences in our models (83%) were higher than the correctly predicted absences (77%). This is the opposite of the usual findings and reflects the difficulty experienced by other studies in predicting the occurrence of rare species (Olden and Jackson 2002). Species presence data are less biased than absence data (Scott et al. 2002).

Our accuracy assessment quantified the performance of fish distribution models against an independent data set, which is the most rigorous method of testing a model (Pearce and Ferrier 2000; Olden et al. 2002; Vaughan and Ormerod 2005). Our correct classification rates (>70% on average) were comparable to those in other studies assessing the accuracy of fish distribution models (Porter et al. 2000; Manel et al. 2001; Olden and Jackson 2001; Filipe et al. 2002; Sylvester 2004; Oakes et al. 2005; Rashleigh et al. 2005). In almost every instance κ indicated moderate to substantial model agreement (>40%), which is slightly better than has been reported in previous studies (Manel et al. 2001; Oakes et al. 2005), in which κ was calculated without weighting the commission error by the DP.

We acknowledge that the DPs were low for the rare species, but they did not have much effect on the accuracy metrics. For example, the hornyhead chub had the largest DP (0.68) but not the largest value for κ or CCR. Several factors influence detection probabilities and modeling accuracy. There are two main types of prediction errors associated with modeling: omission (false negatives) and commission (false positives). By incorporating DPs into our calculations of the accuracy statistics, we rendered our commission error rates more representative of the true error rates. Also, our models contained minimal omissions for all 62 species and the targeted rare species. It is assumed that false negatives are more costly than false positives (Sowa et al. 2005); therefore, these models can be used to identify areas of suitable habitat for rare species across the landscape (Oakes et al. 2005).

Incorporating DPs into accuracy metrics accounted for gear efficiency issues (Hayer 2005). All fish sampling gears are subject to bias depending on the environmental and chemical conditions of the water and the behavioral, ecological, and morphological characteristics of the species present (Paller 1995; Meador et al. 2003; Sylvester 2004). The efficiency of sampling gears may be affected by habitat conditions and species missed as a result. Accounting for heterogeneous DPs among species improved the distribution models in this study.

Several other factors could account for the incorrect classifications in our study, including sampling effort and size, the temporal and spatial variation of fishes, and inaccurate species distribution models. Low values of species occurrence from independent field collections can greatly affect model accuracy by increasing both omission and commission errors (Stockwell and Peterson 2002; Wall et al. 2004) and result in small detection probability. Because most species comprise a small portion of the total fish community (Braaten 1993; Cunningham 1999; Wall et al. 2001), it was difficult to identify true-presence sites. Conversely, it was impossible to identify true-absence sites (Wall et al. 2004).

Sampling effort was variable among the EDUs in this study, being greatest at the East River and Minnesota EDUs. As a result, model accuracy metrics were variable among EDUs, the East River EDU having the highest CCR (mean = 0.828) and κ values (mean = 0.514) for the species of concern. Our data for rare species showed generally increasing predictive accuracy as the number of sample sites increased.

Greater sampling effort in other EDUs may have increased the predictive accuracy of our models.

The variation in the predictive accuracy of models may result from behavioral differences among species that relate to habitat. Habitat has been recognized as a primary gradient for the arrangement and structure of animal communities (Schoener 1974), especially fishes (Schlosser 1987; Arterburn and Berry 2002; Berry et al. 2004; Wall et al. 2004). Individual species preferences and behaviors may be the source of some variation in model accuracy, which may be compounded by other factors, such as seasonal shifts in distribution (Fausch and Bestgen 1996; Filipe et al. 2002). For example, the banded killifish prefers habitats associated with lakes, ponds, and sluggish streams, whereas our sampling focused on wadeable stream sites. Sampling lakes and ponds may have led to the collection of this species and thus increased the predictive accuracy of our models. The finescale dace prefers small lakes in cool, boggy environments associated with springs or beaver dams (Bailey and Allum 1962; Baxter and Stone 1995). Although finescale dace were collected in other habitats, targeting their preferred habitat may have increased the predictive accuracy of our finescale dace model.

We found that stream size and streamflow were the key habitat variables for predicting the presence/absence of most of the fish species of concern, as have other authors (Smith et al. 2002; Wall et al. 2004; Oakes et al. 2005; Sowa et al. 2007). Geographical information system layers are not 100% accurate, and errors accumulate with each layer that is added to the model. Additionally, it is possible that variables not included in our GIS models (e.g., bedrock geology and soils) or factors at finer scales influence fish distributions, so that excluding them would lower the accuracy of our predictions.

Management Applications

Improved knowledge of the distribution of threatened and endangered fishes in South Dakota was an immediate benefit of this study. Our data contributed to the decision to remove the trout-perch and central mudminnow from the list of fish species of concern in 2006. The apparent range of the northern redbelly dace was expanded, and predictions of its presence and that of other species will help guide future surveys. The documented habitat associations will assist with planning watershed management activities, which is the focus of modern riverine management (Williams et al. 1997; Wissmar and Bisson 2003). For example, Wall et al. (2004) used the distribution model for the endangered Topeka shiner *Notropis topeka* to recommend locations for conservation activities.

Validating and assessing the accuracy of fish distribution models has less immediate benefits than those pertaining to fisheries management, but it is an important step in understanding the usefulness of models. The model accuracy assessments in this study give managers high confidence in using models to make decisions about the locations for surveys of habitat management. The models used in this study are available without detection probability correction for fishes in other states in the upper Missouri River basin (Sylvester 2004; Wall et al. 2004). Our results should encourage others to conduct studies incorporating DPs or at least to be aware of the improvements in models that might be realized.

Acknowledgments

The South Dakota Department of Game, Fish and Parks (SDGFP) funded this project as a State Wildlife Grant Project. The South Dakota Gap Analysis Project, particularly the Upper Missouri River Basin Gap Analysis Project provided many of the GIS layers, species locations and modeling techniques. We thank those who allowed us to add their fish collection data to ours for testing our models, specifically S. Freeling, C. Hoagstrom, J. Kral, N. Morey, J. Shearer, R. Sylvester, and S. Thomson. Field technicians who helped collect data for this project are B. Abel, W. Bouska, and D. Doris. We thank landowners who allowed us to study streams on their land. The South Dakota Cooperative Fish and Wildlife Research Unit is jointly sponsored by the SDGFP, Wildlife Management Institute, U.S. Fish and Wildlife Service, SDSU, and U.S. Geological Survey. We appreciate the constructive reviews provided by S. Herrington, K. Schuler, and two anonymous reviewers who helped to improve earlier drafts.

References

Annis, G. M., S. D. Sowa, and D. D. Diamond. 2002. Detailed procedures for classifying stream valley segment types using the National Hydrography Dataset. Missouri Resource Assessment Partnership, University of Missouri, Technical Bulletin 2002-001, Columbia.

Arterburn, J. E., and C. R. Berry, Jr. 2002. Effect of hook style, bait type, and river location on trotline catches of flathead and channel catfish. North American Journal of Fisheries Management 22:573–578.

Bailey, L. L., T. R. Simons, and K. H. Pollock. 2004. Estimating site occupancy and species detection probability parameters for terrestrial salamanders. Ecological Applications 14:692–702.

Bailey, R. M., and M. O. Allum. 1962. Fishes of South Dakota. Miscellaneous Publications Museum of Zoology University of Michigan 119.

Baxter, G. T., and M. D. Stone. 1995. Fishes of Wyoming. Wyoming Game and Fish Department, Cheyenne.

Bayley, P., and J. T. Peterson. 2001. An approach to estimate probability of presence and richness of fish species. Transactions of the American Fisheries Society 130:620–633.

Berry, C. R., M. Wildhaber, and D. Galat. 2004. Fish distribution and abundance. Population structure and habitat use of benthic fishes along the Missouri and lower Yellowstone rivers, volume 3. Cooperative Fish and Wildlife Research Unit, South Dakota State University, Brookings.

Biggs, D., B. de Ville, and E. Suen. 1991. A method of choosing multiway partitions for classification and decision trees. Journal of Applied Statistics 18:49–62.

Boulinier, T., J. Nichols, J. Sauer, J. Hines, and K. Pollock. 1998. Estimating species richness: the importance of heterogeneity in species detectability. Ecology 79:1018–1028.

Braaten, P. J. 1993. The influence of habitat structure and environmental variability on habitat use by fish in the Vermillion River, South Dakota. Master's thesis. South Dakota State University, Brookings.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. Classification and regression trees. Chapman and Hall/CRC Press, New York.

Cassidy, K., E. O. Garton, W. B. Grohn, L. S. Mills, J. M. Scott, K. Williams, and B. Csuti. 1994. Assessing the predictive ability of gap analysis vertebrate distribution maps: gap analysis handbook. Idaho Cooperative Fish and Wildlife Research Unit, Moscow.

Cunningham, G. R. 1999. A survey for the Topeka shiner (*Notropis topeka*) within the Big Sioux, Vermillion, and James river basins in South Dakota. Ecocentrics report prepared for the South Dakota Department of Game, Fish, and Parks, Pierre.

Davis, F. W. 1996. The nature of gap analysis. BioScience 46:74–75.

Dedon, M. F., S. A. Laymon, and R. H. Barrett. 1986. Evaluating models of wildlife–habitat relationships of birds in black oak and mixed conifer habitats. Pages 115–116 *in* J. Verner, M. L. Morrison, and C. J. Ralph, editors. Wildlife 2000: modeling habitat relationships of terrestrial vertebrates. University of Wisconsin Press, Madison.

Edwards, T. C. 1996. Data defensibility and gap analysis. BioScience 46:75–76.

Edwards, T. C., E. T. Deshler, D. Foster, and G. G. Moisen. 1996. Adequacy of wildlife habitat relation models for estimating spatial distributions of terrestrial vertebrates. Conservation Biology 10:263–270.

ESRI (Environmental Systems Research Institute). 1999. ArcView GIS, version 3.2. ESRI, Redlands, California.

ESRI (Environmental Systems Research Institute). 2000. ARC/INFO GIS, version 8.0.2. ESRI, Redlands, California.

Fausch, K. D., and K. R. Bestgen. 1996. Ecology of fishes indigenous to the central and southwestern Great Plains. Pages 131–166 *in* F. Knopf and F. Samson, editors. Ecology and conservation of Great Plains vertebrates. Springer-Verlag, New York.

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

Filipe, A., I. Cowx, and M. Collares-Pereira. 2002. Spatial modeling of freshwater fish in semi-arid river systems: a tool for conservation. River Research and Applications 18:123–136.

Frissell, C. A., J. L. William, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. Environmental Management 10:199–214.

Gaston, K. J. 1994. Rarity. Chapman and Hall, London.

Gu, W., and R. K. Swihart. 2004. Absent or undetected? Effects of nondetection of species occurrence on wildlife–habitat models. Biological Conservation 116:195–203.

Guisan, A., and N. E. Zimmerman. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147–186.

Hayer, C. A., and E. R. Irwin. In press. Influence of gravel mining and other factors on detection probabilities of coastal plain fishes in the Mobile River basin, Alabama. Transactions of the American Fisheries Society.

Hayer, C. A., S. S. Wall, and C. R. Berry. 2006. Evaluation of aquatic gap analysis fish distribution models with emphasis on rare fish species in South Dakota. Report to South Dakota Game, Fish, and Parks, Project T-9, Study 2409, Brookings.

Higgins, J., M. Lammert, M. Bryer, M. DePhilip, and D. Grossman. 1998. Freshwater conservation in the Great Lakes basin: development and application of an aquatic community classification framework. The Nature Conservancy, Chicago.

Higgins, J. V., M. T. Bryer, M. L. Khoury, and T. W. Fitzhugh. 2005. A freshwater classification approach for biodiversity conservation planning. Conservation Biology 19:432–445.

Hoagstrom, C. W., and C. R. Berry, Jr. 2006. Island biogeography of native fish faunas among Great Plains drainage basins: basin scale features influence composition. Pages 221–264 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Landscape influences on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Hoagstrom, C. W., S. S. Wall, J. G. Kral, B. G. Blackwell, and C. R. Berry, Jr. 2007. Zoogeographic patterns and faunal change of South Dakota fishes. Western North American Naturalist 67:161–184.

Hoeting, J. A., M. Leecaster, and D. Bowden. 2000. An improved model for spatially correlated binary responses. Journal of Agricultural, Biological, and Environmental Statistics 5:102–114.

Hogan, E. P. 1995. The geography of South Dakota. Pine Hill Press, Freeman, South Dakota.

Holliday, V. T., J. C. Knox, G. L. Running IV, R. D. Mandel, and C. R. Ferring. 2002. The Central Lowlands and Great Plains. Pages 335–362 *in* A. R. Orme, editor. The physical geography of North America. Oxford University Press, New York.

Jennings, M. D. 2000. Gap analysis: concepts, methods, and recent results. Landscape Ecology 15:5–20.

Karl, J. W., P. J. Heglund, E. O. Garton, J M Scott, N. M. Wright, and R. L. Hutto. 2000. Sensitivity of species habitat-relationship model performance to factors of scale. Ecological Applications 10:1690–1705.

Krohn, W. B. 1996. Predicted vertebrate distributions from gap analysis: considerations in the designs of statewide accuracy assessments. Pages 147–162 *in* J. M. Scott, T. H. Tear, and F. W. Davis, editors. Gap analysis: a landscape approach to biodiversity planning. American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland.

Lammert, M., J. Higgins, D. Grossman, and M. Bryer. 1996. A classification framework for freshwater communities. Proceedings of the Nature Conservancy's Aquatic Community Classification Workshop. The Nature Conservancy, Arlington, Virginia.

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

MacKenzie, D. I., J. D. Nichols, J. Hines, M. G. Knutson, and A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 848:2200–2207.

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

MacKenzie, D. I., J. D. Nichols, N. Sutton, K. Kawanishi, and L. L. Bailey. 2005. Improving inferences in population studies of rare species that are detected imperfectly. Ecology 86:1101–1113.

Manel, S., H. C. Williams, and S. J. Ormerod. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38:921–931.

Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possingham. 2005. Zero tolerance ecology: improving ecological inference by modeling the source of zero observations. Ecology Letters 8:1235–1246.

McArdle, B. H. 1990. When are rare species not there? Oikos 57:276–277.

McDonald, L. L. 2004. Sampling rare populations. Pages 11–42 *in* W. L. Thompson, editor. Sampling rare or elusive species. Island Press, Washington, D.C.

Meade, R. H., T. R. Yuzyk, and T. J. Day. 1990. Movement and storage of sediment in rivers of the United States and Canada. Pages 255–280 *in* M. G. Wolman and H. C. Riggs, editors. Surface water hydrology: the geology of North America, volume O-1. Geological Society of America, Boulder, Colorado.

Meador, M., J. McIntyre, and K. Pollock. 2003. Assessing the efficacy of single-pass backpack electrofishing to characterize fish community structure. Transactions of the American Fisheries Society 132:39–46.

Nichols, J. D., T. Boulinier, J. E. Hines, K. H. Pollock, and J. R. Sauer. 1998a. Estimating rates of local species extinction, colonization, and turnover in animal communities. Ecological Applications 8:1213–1225.

Nichols, J. D., T. Boulinier, J. E. Hines, K. H. Pollock, and J. R. Sauer. 1998b. Inference methods for spatial variation in species richness and community composition when not all species are detected. Conservation Biology 12:1390–1398.

Oakes, R. M., K. B. Gido, J. A. Falke, J. D. Olden, and B. L. Brock. 2005. Modeling of stream fishes in the Great Plains, USA. Ecology of Freshwater Fish 14:361–374.

Olden, J. D., and D. A. Jackson. 2001. Fish–habitat relationships in lakes: gaining predictive and explanatory insight by using artificial neural networks. Transactions of the American Fisheries Society 130:878–897.

Olden, J. D., and D. A. Jackson. 2002. A comparison of statistical approaches for modeling fish species distributions. Freshwater Biology 47:1976–1995.

Olden, J. D., D. A. Jackson, and P. R. Peres-Neto. 2002. Predictive models of fish species distributions: a note on proper validation and chance predictions. Transactions of the American Fisheries Society 131:329–336.

Omernik, J. M. 1987. Ecoregions of the conterminous United States [map (scale 1:7,500,000)]. Annals of the Association of American Geographers 77:118–125.

Paller, M. 1995. Relationships among number of fish species sampled, reach length surveyed, and sampling effort in South Carolina coastal plain streams. North American Journal of Fisheries Management 15:110–120.

Pearce, J., and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling 133:225–245.

Pollock, K. H., J. D. Nichols, T. R. Simons, G. L. Farnsworth, L. L. Bailey, and J. R. Sauer. 2002. Large-scale wildlife monitoring studies: statistical methods for design and analysis. Environmetrics 13:1–15.

Porter, M. S., J. Rosenfeld, and E. A. Parkinson. 2000. Predictive models of fish species distribution in the Blackwater drainage, British Columbia. North American Journal of Fisheries Management 20:349–359.

Rashleigh, B., R. Parmar, J. M. Johnston, and M. C. Barber. 2005. Predictive habitat models for the occurrence of stream fishes in the mid-Atlantic Highlands. North American Journal of Fisheries Management 25:1353–1366.

Rushton, S. P., S. J. Omerod, and G. Kerby. 2004. New paradigms for modeling species distributions? Journal of Applied Ecology 41:193–200.

Schlosser, I. J. 1987. The role of predation in age-related and size-related habitat use by stream fishes. Ecology 68:651–659.

Schmidt, B. R. 2003. Count data, detection probabilities, and the demography, dynamics, distribution, and decline of amphibians. Comptes Rendus Biologies 326:119–124.

Schoener, T. W. 1974. Resource partitioning in ecological communities. Science 185:27–39.

Scott, J. M., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D. Erchia, T. C. Edwards, J. Ulliman, and G. Wright. 1993. Gap analysis: a geographic approach to protection of biological diversity. Wildlife Monographs 123.

Scott, J. M., P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Rafael, W. A. Wall, and F. B. Samson, editors. 2002. Predicting species occurrences: issues of accuracy and scale. Island Press, Washington, D.C.

Scott, J. M., M. Jennings, R. G. Wright, and B. Csuti. 1996. Landscape approaches to mapping biodiversity. BioScience 46:77–78.

SDGFP (South Dakota Department of Game, Fish, and Parks). 2007. Rare, threatened or endangered animals tracked by the South Dakota Natural Heritage Program. Available: www.sdgfp.info/wildlife/Diversity/RareAnimal.htm. (October 2007).

Short, H. L., and J. B. Hestbeck. 1995. National biotic resource inventories and gap analysis. BioScience 45:535–539.

Smith, V. J., J. A. Jenks, C. R. Berry, Jr., C. J. Kopplin, and D. D. Fecske. 2002. South Dakota Gap Analysis Project, final report. U.S. Geological Survey, Reston, Virginia.

Sowa, S. P., G. Annis, M. E. Morey, and D. D. Diamond. 2007. A gap analysis and comprehensive conservation strategy for riverine ecosystems of Missouri. Ecological Monographs 77:301–334.

Sowa, S. P., D. D. Diamond, R. Abbitt, G. M. Annis, T. Gordon, M. E. Morey, G. R. Sorenson, and D. True. 2005. A gap analysis for riverine ecosystems of Missouri. Final Report to the U.S. Geological Survey, National Gap Analysis Program. Available: www.cerc.usgs.gov/morap/projects.asp. (June 2008).

SPSS. 2001. AnswerTree 3.0 user's guide. SPSS, Chicago.

Stockwell, D. R., and A. T. Peterson. 2002. Effects of sample size on accuracy of species distribution models. Ecological Modelling 148:1–13.

Sylvester, R. M. 2004. Upper Missouri River basin aquatic gap fish distribution model accuracy assessment and white sucker, *Catostomus commersonii*, population characteristics in the upper Missouri River basin. Master's thesis. South Dakota State University, Brookings.

Thornbury, W. D. 1965. Regional geomorphology of the United States. Wiley, New York.

TNC–FWI (The Nature Conservancy–Freshwater Initiative). 2000. GIS tools for aquatic macrohabitat classification. Available: www.freshwaters.org/ccwp/home.html. (February 2000).

USGS (U.S. Geological Survey). 2002. National Hydrography Dataset. Available: nhd.usgs.gov/index.html. (April 2007).

Vaughan, I. P., and S. J. Ormerod. 2005. The continuing challenges of testing species distribution models. Journal of Applied Ecology 42:720–730.

Wall, S. S., C. R. Berry, Jr., C. M. Blausey, J. A. Jenks, and C. J. Kopplin. 2004. Fish habitat modeling for gap analysis to conserve the endangered Topeka shiner (*Notropis topeka*). Canadian Journal of Fisheries and Aquatic Sciences 61:954–973.

Wall, S. S., C. M. Blausey, J. A. Jenks, and C. R. Berry, Jr. 2001. Topeka shiner (*Notropis topeka*) population status and habitat conditions in South Dakota streams. Department of Wildlife and Fisheries Sciences, South Dakota State University, Brookings.

Wall, S. S., V. J. Smith, R. M. Sylvester, J. A. Jenks, C. R. Berry, and C. J. Kopplin. 2005. The Upper Missouri River Basin Aquatic Gap Analysis Project, draft report. South Dakota Cooperative Fish and Wildlife Research Unit, South Dakota State University, Brookings.

Williams, J. E., C. Wood, and M. Dombeck. 1997. Watershed restoration: principles and practices. American Fisheries Society, Bethesda, Maryland.

Winter, T. C., and M.-K. Woo. 1990. Hydrology of lakes and wetlands. Pages 159–187 *in* M. G. Wolman and H. C. Riggs, editors. Surface water hydrology: the geology of North America, volume O-1. Geological Society of America, Boulder, Colorado.

Wintle, B. A., M. A. McCarthy, K. M. Parris, and M. A. Burgman. 2004. Precision and bias of methods for estimating point survey detection probabilities. Ecological Applications 14:703–712.

Wissmar, R. C., and P. A. Bisson, editors. 2003. Strategies for restoring river ecosystems: sources of variability and uncertainty in natural and management systems. American Fisheries Society, Bethesda, Maryland.