Fisheries Research 147 (2013) 359–369

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Improving macroscopic maturity determination in a pre-spawning flatfish through predictive modeling and whole mount methods

Stephen R. Midway[a], J. Wilson White[a], William Roumillat[b], Christopher Batsavage[c,*], Frederick S. Scharf[a]

[a] Department of Biology and Marine Biology, University of North Carolina Wilmington, Wilmington, NC 28403, USA
[b] Marine Resources Research Institute, South Carolina Department of Natural Resources, Charleston, SC 29412, USA
[c] North Carolina Division of Marine Fisheries, Morehead City, NC 28557, USA

ARTICLE INFO

Article history:
Received 13 November 2012
Received in revised form 16 July 2013
Accepted 18 July 2013

Keywords:
Paralichthys lethostigma
AIC
Cross-validation
Hindcasting
Gonadosomatic index

ABSTRACT

Accurate maturity schedules are essential for informed management of many fishery resources. Although histological methods are generally acknowledged as the best approach to correctly assign maturity status of individual fish, the methods can be expensive and time consuming. We developed and tested a set of multivariable models to predict maturity of southern flounder, a valuable flatfish occupying estuarine and coastal systems in the southeastern US. We also evaluated the potential for whole mount methods to validate maturity assignments and help discriminate transitional oocyte stages. Lastly, we used one of the better performing models to conduct retrospective analysis of variability in southern flounder maturity schedules. Several models performed well in predicting southern flounder maturity; nearly half of the models we tested achieved ≥85% prediction success. We noted that the gonadosomatic index (GSI) was included in most of the higher performing models and, by itself, was a strong predictor of maturity for southern flounder. The addition of novel quantitative predictors, such as gonad color and dimensions, pushed model success above 90% in many cases. Whole mount methods showed a high level of agreement with histological methods, and should be investigated as an inexpensive alternative for validating maturity assignments. Retrospective analysis revealed the potential for interannual fluctuations in L_{50} of 2–5 cm for southern flounder, which can impact yearly estimates of SSB and target harvest rates. Multivariable predictive models using routinely collected fishery biological data can provide reliable estimates of fish maturity and, when coupled with whole mount methods, should represent an improvement of traditional macroscopic maturity assignment.

1. Introduction

Understanding temporal variability in the timing of maturity is essential for successful management of exploited fish stocks, yet maturity schedules are often incomplete or prone to error (Lowerre-Barbieri et al., 2011a). For many fishes, maturity assignments have traditionally been based on visual inspection of gross (macroscopic) features of whole gonads (West, 1990). However, macroscopic staging can be inaccurate since it is generally restricted to coarse and often subjective measures of gonad size, shape, or color that do not necessarily correspond with oocyte-level development (e.g., Vitale et al., 2006; Costa, 2009; Ferreri et al., 2009; McPherson et al., 2011; Midway and Scharf, 2012). Despite its limitations, macroscopic staging remains a common approach for assigning maturity status since it is inexpensive and relatively easy to complete. Alternatively, histological examination of gonadal tissue is widely considered to be the benchmark in fish reproductive biology as it provides oocyte-level information and a high degree of accuracy in distinguishing between immature and mature individuals (Hunter and Macewicz, 1985; Murua and Saborido-Rey, 2003). The major drawback to histology is that it is resource intensive, requiring considerable time and expense, and specialized training.

In some cases, accurate maturity assignments have been obtained using simple and inexpensive methods instead of histology (West, 1990; Neidig et al., 2000). Specifically, two alternative approaches that have promise include multivariable predictive modeling and whole mount procedures. Statistical models to predict maturity status for individual fish can take a number of forms, but recent models have generally taken advantage of routinely collected biological (e.g., length, weight, age, etc.) and environmental (e.g., location and date of capture, physical habitat attributes, etc.) variables to increase predictive success beyond simple macroscopic

* Corresponding author. Tel.: +1 9197935386; fax: +1 9109624066.
E-mail addresses: steve.midway@gmail.com (S.R. Midway), whitejw@uncw.edu (J.W. White), roumillatb@dnr.sc.gov (W. Roumillat), chris.batsavage@ncdenr.gov (C. Batsavage), scharff@uncw.edu (F.S. Scharf).

0165-7836/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.fishres.2013.07.010

staging criteria. For example, Vitale et al. (2006) used a regression tree model to determine that gonadosomatic (GSI) and hepatosomatic (HSI) indices could correctly predict maturity of Kattegat Atlantic cod (*Gadus morhua*) for 95% of the individuals tested. Similarly, Schill et al. (2010) used environmental variables to generate maturity models with prediction success rates between 77 and 89% for Idaho redband trout (*Oncorhynchus mykiss gairdneri*). The ability to predict maturity status with high rates of success by supplementing macroscopic staging with information gained from multiple routinely collected variables could lessen the need for regular histological analysis.

Whole mounting of oocytes (also referred to as wet or squash mounts) is a simplified method to identify oocyte stages that requires only placing a sample of fresh or preserved ovarian tissue on a glass slide, compressing the sample with a cover slip, and then viewing using a stereo or compound microscope. Whole mount methods have been applied more commonly as part of fecundity studies (e.g., Thorsen and Kjesbu, 2001; Witthames et al., 2009) than for the development of maturity schedules; however, staging of oocytes from whole mounts has been successfully validated in cases where it was attempted (Kjesbu, 1991; Neidig et al., 2000). More specifically, both Forberg (1983) and West (1990) reported high agreement between whole mount and histological methods to stage oocytes of capelin (*Mallotus villosus villosus*) and brownstripe red snapper (*Lutjanus vittus*), respectively. West (1990) further suggested that, if validated, whole mounts can be more efficient than histology for generating maturity schedules. Similar to simple GSI calculations, whole mounting is rapid and inexpensive; however it can yield microscopically detailed information for some species, indicating its promise as a potential alternative to more extensive histological approaches.

The ability to produce rapid and accurate maturity schedules on a continual basis should increase our understanding of spatial and temporal variation in maturity and lead to more effective management (Morgan, 2008). Fish maturity schedules can vary for numerous reasons, including density-dependent responses to fluctuations in population abundance (Morgan and Colbourne, 1999; Hutchings and Baum, 2005), environmental variation (e.g., temperature; Dhillon and Fox, 2004; Tobin and Wright, 2011), and selective mortality (Law and Grey, 1989). Because changes in maturity are thought to be influenced more by changes in abundance and demographics (i.e., selective removal of individuals through fishing) and less by natural variation (Walsh and Morgan, 1999), exploited species are more likely to display temporal variability in maturation schedules. The ability to monitor temporal shifts in age- and size-specific maturity has the potential to serve as a warning of overfishing (Morgan, 2008), thus quantifying and understanding causes of variation in reproductive timing should be a research priority.

In this study, we attempt to improve upon existing macroscopic maturity staging for female southern flounder (*Paralicthys lethostigma*), an economically valuable flatfish in coastal waters of the U.S. South Atlantic and Gulf of Mexico. The latest assessment of the population harvested in North Carolina waters concluded that the stock was overfished (spawning stock biomass [SSB] below the threshold level; Takade-Heumaker and Batsavage, 2009). Recent findings (Midway and Scharf, 2012) also indicate that the macroscopic maturity classification system used in the assessment likely leads to overestimation of SSB and overly optimistic biological reference points. Overfished status combined with lower SSB than previously thought highlights the need for accurate maturity information so that future assessments and management decisions for southern flounder can be conducted with confidence.

Misclassification of southern flounder maturity likely stems from two sources, which together were suspected to be the cause of low agreement between macroscopic and histological maturity assignments for southern flounder in early stages of reproductive development (Midway and Scharf, 2012). The first source of misclassification is the application of mostly descriptive macroscopic staging criteria that have not been thoroughly validated. Second, southern flounder have historically been inaccessible during spawning, which is presumed to take place in deep offshore waters, meaning that most fish are collected during the months immediately preceding spawning and thus, only possess early oocyte developmental stages. Given the resources necessary to complete histological analyses, it is unlikely that they will become routine for southern flounder or other similar species managed by state and federal agencies. However, as shown for other teleosts, routinely collected biological variables (e.g., gonadosomatic indices, body mass, condition indices, age, date of capture) can enhance maturity predictions and are less time and cost intensive. Thus, there is strong potential to improve maturity predictions for southern flounder and thereby enhance future stock assessments and management decisions. To address this possibility, our objectives were: (1) to combine existing and novel macroscopic characters with additional biological variables to develop and evaluate multivariable models for maturity prediction; (2) to assess the potential for using whole mount methods to generate maturity schedules; and (3) to demonstrate the utility of multivariable predictive models by quantifying historic variability in size-based maturity of North Carolina southern flounder.

Table 1
Stages used to assess maturity in North Carolina southern flounder. Each stage and the description of its macroscopic features follow guides used by the North Carolina Division of Marine Fisheries. The most advanced oocyte stages are based upon histological examination of female gonadal sections and follow Brown-Peterson et al. (2011). CA = cortical alveolar; OM = oocyte maturation; PG = primary growth; POF = postovulatory follicle complex; Vtg1 = primary vitellogenesis; Vtg2 = secondary vitellogenesis; Vtg3 = tertiary vitellogenesis.

Macroscopic stage	Macroscopic features	Histologically most advanced stage oocyte
Immature	Ovaries small and thin—no oocytes visible	PG
Developing	Ovaries rotund, yellowish-orange and turgid	CA, Vtg1, Vtg2
Fully developed	Same as developing, but with oocytes visible	Vtg3
Ripe (running)	Ovaries large and soft with many large, free-flowing (with slight pressure) hydrated oocytes	Vtg3, OM, POF
Spent	Ovaries small and bloodshot; few hydrated oocytes, if any	POF, few Vtg
Resting	Ovaries small, flaccid, translucent with no visible oocytes	PG

2. Materials and methods

2.1. Fish collection and measured variables

Southern flounder were collected during the fall (Oct–Dec) of 2009 and 2010. A complete description of collection methods and data retrieval is contained in Midway and Scharf (2012). Briefly, fish were collected from both fishery-dependent and -independent sources, kept on ice, and returned to the laboratory for processing. Variables measured for all fish included total length (TL) in mm, weight (g), age (y), gonadosomatic index (GSI = [gonad weight/(body weight − gonad weight)] × 100), ordinal date of capture, and assigned macroscopic maturity stage (stages are those used by the North Carolina Division of Marine Fisheries [NCDMF] and are described in Table 1). Fish were randomly selected for histological analysis within each year (2009

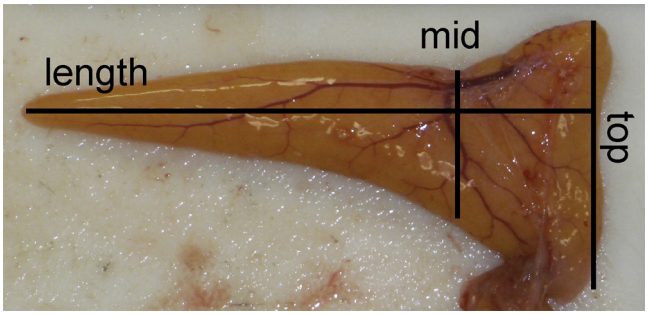

Fig. 1. Initial measurements used in ovary morphology descriptors: girth = mid/length; top thickness = mid/top; elongation = length/top.

n = 199; 2010 n = 225) after stratifying the data by collection month and region. For both years, fish TL was approximately normal with the mean ± SD = 388.4 ± 52.5 mm and 385.5 ± 51.4 mm during 2009 and 2010, respectively. Southern flounder display strong sexual dimorphism in size, and the fishery targets larger females. Consequently, nearly all fish we collected were females, which were the focus of our analyses.

In addition to the variables measured for all fish, six additional variables were measured for a subset of individuals. These included two measures of gonad coloration (i.e., mean color density, mean red density), three measures of ovary dimension (i.e., ovary girth, top thickness and elongation) and the longitude of the capture location. We hypothesized that quantifying color variation among ovaries would be a useful predictor of maturity since most macroscopic staging guides include descriptions of color, including guides for southern flounder (Wenner et al., 1990). Additionally, Peer et al. (2012) recently used measurements of gonad color to successfully discriminate several oocyte stages of striped bass (*Morone saxatilis*). During macroscopic staging, we used a Fujifilm FinePix 8.2 megapixel digital camera to take digital images of whole ovaries for a subset (n = 195) of fish. The camera was fixed to a rigid frame and positioned approximately 305 mm above the gonad, and no flash was used. Images were uniformly white balanced using Image Pro software (Media Cybernetics, Bethesda, MD, USA) to standardize residual variation, although photographic location, focal length, and camera settings were consistent for all images, which should have minimized any color-balancing issues. Next, we identified a representative region of the ovary (approximately 10% of the area) which was selected to exclude connective tissue, blemishes, or other imperfections that were not representative of the ovary color. This sample was then used for estimation of mean color density and mean red density, both of which are objective, digital measures of saturation. Due to observations of changing ovary morphology during maturation—e.g., lobe thickening and elongation—we also hypothesized that ratios of ovary size (length/width) might provide a quantitative measure that would enhance maturity prediction. During ovary processing, one lobe was randomly selected to obtain measurements of length, width at the top (top width), and width in the middle (mid width; defined as the notch where the stout upper part of the gonad meets the tapered lobe) (see Fig. 1). From these measurements, three ratios were created, including girth (mid width/length), top thickness (mid width/top width) and elongation (length/top width), and included as maturity predictors for a subset (n = 230) of fish. Lastly, longitude was hypothesized to be a possible predictor of maturity since some of North Carolina's larger bays (e.g., Albemarle and Pamlico Sounds) extend across large east–west distances (>1° of longitude) from riverine nursery habitats to ocean inlets. Fish captured at eastern longitudes were closer to ocean inlets, indicating a potential readiness to initiate an offshore spawning migration. Thus, we speculated that these individuals might possess more advanced oocytes, and we examined longitude as a maturity predictor for 73 individuals collected from these Sounds.

2.2. Model construction and selection

We used logistic regression models to examine a total of eight macroscopic variables for their performance in predicting maturity of southern flounder. All of the fish used in this analysis had been examined histologically and categorized as sexually mature or immature based on the most advanced oocyte stages identified. Based on oocyte stages described in Grier et al. (2009), we used the presence of cortical alveolar oocytes (or more advanced oocyte stages) to define a mature fish, assuming that fish which had initiated oocyte development would complete the process and spawn in the upcoming spawning season (Murua and Saborido-Rey, 2003; Brown-Peterson et al., 2011; Lowerre-Barbieri et al., 2011b). These histological maturity assignments were assumed to reflect the 'true' maturity status, and the performance of predictive models in determining maturity was judged relative to histological assignments. Our first set of models tested six primary variables (assigned macroscopic stage, GSI, TL, weight, date of capture, and age) as well as the interactions [GSI × date of capture] and [TL × date of capture] to determine which combination of those variables produced the most accurate maturity assignment. The choice of these two specific interactions was based on the likely time required to complete both oocyte development and offshore migration, and thus the hypothesis that a female individual would need to initiate oocyte development by a certain time of year in order to participate in the upcoming spawning season. Therefore, we surmised that achieving some threshold of somatic growth (TL) or gonadal development (GSI) by a specific time of year might represent a good predictor of maturity. We created a set of 255 models that included all possible combinations of the predictor variables. Multivariable logistic regression models with binomial error distribution and logit link function were then fitted using the glm function in R version 2.15 (R Development Core Team, 2012).

Model performance was first assessed using Akaike's information criterion (AIC). Model selection based on AIC is commonly applied in ecological studies in order to compromise between goodness-of-fit and model complexity (Burnham and Anderson, 2002; Symonds and Moussalli, 2011). Model complexity refers to the number of parameters, with an increased number of parameters often resulting in the model closely fitting the training data; however, this complexity may or may not be useful for predicting new observations. We used AICc—which converges on AIC with increasing sample size—due to low sample size relative to the number of model parameters (Burnham and Anderson, 2002). In our analysis, all models with ΔAICc ≤ 2 (where ΔAICc is measure of each model AIC relative to the best model AIC) were considered plausible and well supported by the data. We also considered models with ΔAICc values as high as 6 and used model probability weights (w_i; defined as the relative weight of evidence supporting a model relative to all other models within the set) to identify a subset of best models (Burnham et al., 2011).

A second framework for evaluating model performance included the use of cross validation approaches, in which a subset of the data are left out of the model and used instead for model testing (Shao, 1993). The strength of cross validation is that it allows for extra-sample inference (i.e., inference or prediction of unknown data—an obvious primary goal of any maturity assignment model), while AICc rankings are based solely on the specific data used to build the models. We first conducted a 2-fold cross validation, in which models were built (i.e., trained) using data from a single year (2009 or 2010) and then tested on the other year. We also conducted jackknife reclassification (i.e., leave-one-out cross validation), in which models were built excluding one observation and

the maturity status of that observation was then predicted by the model. For each fish in the data set, the absolute value of the difference between the predicted probability of maturity and actual maturity (either 0 or 1) was calculated. The model was deemed successful for a fish if this difference was <0.5. For example, if a fish were determined to be mature based on histology, a model prediction of 80% maturity probability would be deemed a model success, whereas a model prediction of 40% maturity probability would be deemed a model failure. The overall performance of cross validation for each model was expressed as the proportion of model successes, and was compared among candidate models.

Both AIC and cross validation have strengths for model evaluation. AIC scores are based on the likelihood of the model given a specific dataset, and will tend to be highest for models that minimize residual error, whereas cross validation success is based purely on correct classification of independent data, which is often a goal in fisheries science. Because our modeling goal was to maximize predictive power while minimizing variables, the use of the two model selection methods provided sets of optimal models based on each selection criteria that could be contrasted. In addition, by understanding the performance of each method with respect to their differences, we would be able to assess a particular model's utility based on the available data. We were concerned that cross validation success might be high due to the cutoff of 50%, whereas the likelihood-based AIC approach would provide a better probabilistic fit. In order to address this concern, we examined the distribution of prediction deviations for best models identified by each selection process. These deviations were defined as the absolute value of the difference between the model-predicted probability of maturity and the histologically determined maturity for a given fish. Since AICc rankings will be higher for models with lower residual error given a fixed number of parameters, we were interested in exploring how models with relatively poor AICc rankings, but high cross validation success, compared to the best AICc models.

Finally, in order to more closely examine the deviation patterns for some of the most successful models, we determined whether prediction failures consisted of many of the same individuals that may have possessed specific sets of traits. We also examined the distribution of failure types to determine if any bias existed, such as most failures resulting from immature model predictions for mature fish, or vice versa. Lastly, we quantified the similarity of specific failures between models. To do this, we calculated multiple pairwise comparisons among the specific individuals comprising the failures from a select group of models and used Sørensen's similarity coefficient (Sørensen, 1948) to characterize specific failures that occurred in both models. Sørensen's similarity coefficient is an overlap index used to measure the similarity of two samples. Values range from 0 and 1, where 0 indicates no overlap and 1 indicates complete overlap.

2.3. Hindcasting variation in maturity schedules

To illustrate the utility of the predictive models for quantifying temporal variation in the timing of maturity, we applied a well-performing (i.e., high jackknife reclassification success) model to an 18-year data set (1991–2008) provided by the NCDMF to estimate historic interannual variability in southern flounder maturity schedules. The data set consisted of fish collected as part of fishery-dependent and -independent sampling programs, collectively known as Program 930 (see Takade-Heumaker and Batsavage, 2009 for additional programmatic details). From the data set, we selected fish for analysis that were captured during the same months (Oct–Dec) as those used to build our models. Based on our model selection results, we identified a single model for use in hindcasting that included TL, weight, and the [GSI × date of capture] interaction as variables. This model demonstrated the highest jackknife reclassification success among models that did not include any novel predictors and was also simple enough that each of the variables was measured for a sufficient number of fish during most of the years in the NCDMF data set. Use of the most supported AICc models, which included 7+ variables, would have resulted in a large number of years in the historical data set with insufficient data. The simpler three-variable logistic model was fitted to generate maturity ogives for all years with sufficient data. Estimates of L_{50} (the length at which 50% of individuals were mature) were calculated by solving the equation for the logistic model for particular values of the predictors:

$$p = \frac{1}{1 + \exp(\boldsymbol{\beta}\mathbf{X})}$$

where p = probability of maturity, $\mathbf{X}$ is a $k \times 1$ vector of predictors (including length) and $\boldsymbol{\beta}$ is a $1 \times k$ vector of coefficients. Years with sufficient data were restricted to those containing fish with enough contrast in TL to model a full maturity ogive (having both a lower and upper asymptote); years generating only partial ogives were omitted from the analysis. Retained years ($n = 9$) nearly always had sample sizes >100 fish (mean $n = 134$), whereas omitted years ($n = 9$) had considerably fewer samples (mean $n = 22$). We also hindcasted southern flounder maturity ogives using a model with only a single predictor—the traditional macroscopic stages that were reported in the data set for the same subset of years. This model would be representative of maturity schedule estimation based only on macroscopic stage assignment.

2.4. Examination of ovary whole mounts

Preserved ovarian tissue that remained after removing sections for histological analysis was used to examine the potential use of whole mounts to assign maturity of southern flounder. Following Neidig et al. (2000), we used a scalpel (size 11) to scrape a small amount of oocytes from an incision in one of the gonad lobes. Typically, samples had been in fixative for >2 years, so the agitation method outlined by Lowerre-Barbieri and Barbieri (1993) had limited success. In addition, this past method was developed to separate large quantities of oocytes for fecundity estimation while whole mounting generally only requires a small sample. Once the sample was obtained, it was placed on a 75 mm × 25 mm glass microscope slide, a drop of water was added, a cover slip was placed on the sample, and then viewed on a compound microscope (Leica DME) typically between 100 and 200× magnification. To identify repeatable features specific to oocyte stages, we initially examined a training set of 10 whole mount samples (representing various oocyte stages) while directly comparing them with histological preparations for the same fish. We then staged 105 whole mount samples blindly using the oocyte descriptions developed from examination of the training set.

3. Results

3.1. Macroscopic model performance

Among the 255 logistic models that we tested, only three models achieved a $\Delta\text{AICc} \leq 2$ (including one model with a ΔAICc just slightly >2) (Table 2). Together, these three models received a high level of support, with a combined relative probability weight (w_i) of 0.52 (Table 2). All three models included six or more predictor variables, indicating that for maturity prediction, more complex models received greater AICc support than simple models.

When cross validation results were compared to AICc rankings, there was a relatively strong correlation between model predictive performance and AICc support (Spearman's $\rho = -0.677$;

Table 2
The ten most supported models according to AICc ranking and their associated cross validation performance. Model numbers indicate the predictors included in each model: 1 = macroscopic stage; 2 = GSI; 3 = TL (mm); 4 = weight (g); 5 = ordinal date of capture; 6 = age; 7 = [TL × ordinal date of capture]; 8 = [GSI × ordinal date of capture]. K = the total number of terms included in the model, which equals the number of predictors plus the intercept. ΔAICc is the difference between each model and the best performing model (with models in bold text for ΔAICc approximately ≤ 2), and w_i is the model probability weight. Subscripts on the 2-way columns indicate which year's data was used to build the model (i.e., performance was assessed on the opposite year), and the Jackknife column represents the reclassification success of the full data set when evaluated with leave-one-out cross validation.

Model	K	AIC_c	ΔAIC_c	w_i	2-way$_{2009}$	2-way$_{2010}$	Jackknife
1234578	**8**	**310.66**	**0.00**	**0.29**	**84.00%**	**81.91%**	**84.20%**
234578	**7**	**312.43**	**1.77**	**0.12**	**85.33%**	**82.91%**	**85.61%**
12345678	**9**	**312.69**	**2.03**	**0.11**	**84.44%**	**80.90%**	**83.96%**
123457	7	313.04	2.38	0.09	83.56%	80.90%	84.43%
2345678	8	314.47	3.81	0.04	84.89%	81.41%	85.38%
23457	6	314.68	4.02	0.04	84.00%	80.90%	87.03%
1234567	8	315.11	4.45	0.03	83.11%	80.90%	84.20%
134578	7	315.13	4.47	0.03	84.00%	79.90%	85.38%
123478	7	315.90	5.24	0.02	84.44%	84.42%	84.43%
1234	5	316.55	5.89	0.02	83.56%	82.91%	85.14%

Fig. 2). However, this correlation weakened to −0.245 when considering only the better models—those with AIC scores <350 and jackknife reclassification success rates >83% (which comprised about three-fourths of all models). During jackknife reclassification, many of the models that we tested achieved equally high predictive success (Fig. 3). Specifically, 208 models (or 82% of all models tested) successfully predicted maturity for >80% of the fish, and 117 models (46% of all models tested) successfully predicted maturity for >85% of the fish. Additionally, when models were grouped by complexity (i.e., the number of predictor variables), at least one model within each group had a jackknife reclassification success rate as high as 84%, including success rates of 85% and 86% for the least complex models that included only one or two predictor variables, respectively (Table 3). For the 2-fold cross validation results, average success rates were very similar for both years: mean ± SD = 81.4 ± 0.05% and 81.6 ± 0.04% for models trained on 2009 and 2010 data, respectively. Interestingly, the model receiving the most AICc support (model 1234578) and the model with the best jackknife reclassification performance (model 348 was the simplest model to achieve 87% prediction success) differed markedly in the number of predictor variables they included (Table 4), although model 1234578 was also well-supported in cross-validation (84%).

When we explored patterns in the deviation of model predictions from histological maturity assignments, the models with relatively poor AICc rankings but high cross validation success had a similar distribution of deviations as did the most supported AICc models (Fig. 4). There was only a slight increase in median deviation, from 0.11 for the most supported AICc models to 0.15 for the models with relatively poor AICc rankings but high cross validation success. Model failures occurred consistently for individuals with intermediate GSI values (mean GSI ± SD = 0.51, 0.52, 0.52, and 0.53 ± 0.28 for misspecified individuals by the four models included in Fig. 4), an indication that GSI contrast was a strong predictor of maturity. Indeed, the GSI-only model did achieve high cross validation success. Additionally, the failures by the models presented in Fig. 4 were mostly in the same direction; the models incorrectly predicted mature fish to be immature

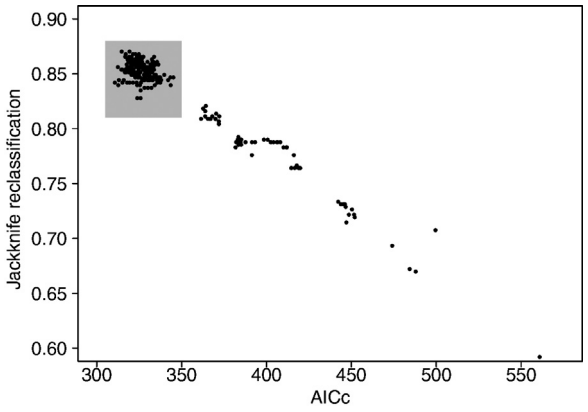

Fig. 2. Scatterplot of all model AICc values and Jackknife reclassification success rates. Model ranking methods were well correlated (Spearman's $\rho = -0.677$); however, the correlation was less strong when considering only the best models for each criteria (the points contained in the gray box; Spearman's $\rho = -0.245$).

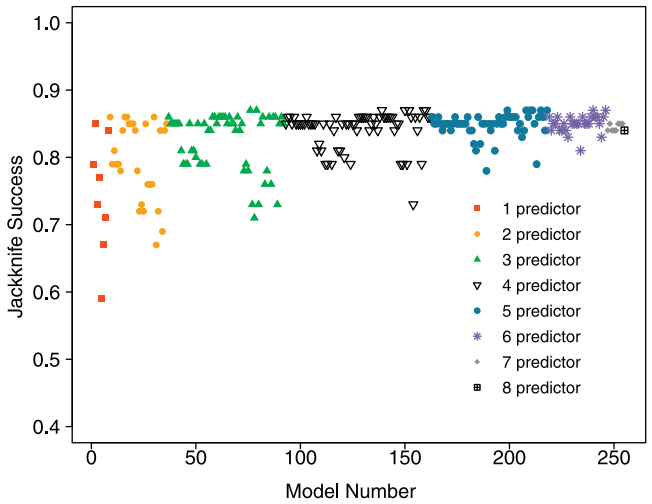

Fig. 3. Leave-one-out cross validation results for models constructed using data from both years combined ($n = 255$ models). At least one model representing each level of model complexity (i.e., inclusion of a specified number of predictor variables) performed well (≥84% classification success).

Table 3
Jackknife reclassification rankings for the best performing model at each level of model complexity, where complexity is defined at the number of predictor terms. Each model below achieved the highest cross validation success among all models that included a specified number (1–8) of predictor variables. Column headings follow those in Table 2.

Model	K	AIC_c	ΔAIC_c	w_i	2-way$_{2009}$	2-way$_{2010}$	Jackknife
2	2	333.5	22.9	0.00	83.11%	87.44%	84.67%
23	3	329.9	19.2	0.00	84.00%	87.44%	86.08%
348	4	319.3	8.6	0.00	84.89%	88.44%	87.03%
3468	5	321.3	10.7	0.00	84.89%	85.93%	86.79%
23457	6	314.7	4.0	0.04	84.00%	80.90%	87.03%
234567	7	316.7	6.1	0.01	84.00%	81.41%	86.56%
1345678	8	317.2	6.5	0.01	83.56%	79.90%	85.14%
12345678	9	312.7	2.0	0.11	84.44%	80.90%	83.96%

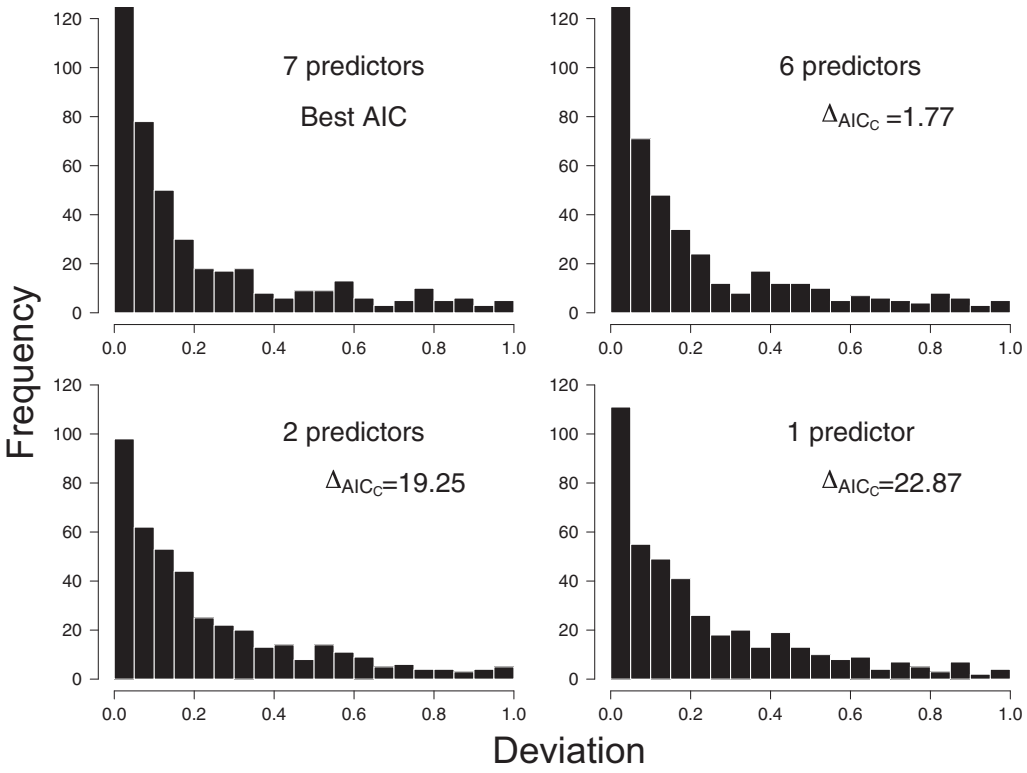

Fig. 4. Histograms of individual deviations (absolute value of the difference between model probability of maturity and histologically determined maturity) for four models. The top two panels include results for the most supported (model 1234578) and second most supported (model 234578) models based on AICc rankings. The bottom two panels include results for models receiving poor AICc support, but achieving high cross validation success (model 23 on left and model 2 on right).

Table 4
Logistic regression coefficients, standard errors (SE), and *P*-values for each term in the model with the best jackknife reclassification (model 348) and the model receiving the most AICc support (model 1234578).

Model	Variable	Coefficient	SE	*P*
348	Intercept	3.0063	3.1987	0.347
	TL (mm)	−0.0325	0.0117	0.005
	Weight (g)	0.0074	0.0019	<0.001
	GSI × ordinal date	0.0223	0.0032	<0.001
1234578	Intercept	49.5612	17.4277	0.004
	Macroscopic stage	0.6358	0.3200	0.047
	GSI	35.8093	13.3284	0.007
	TL (mm)	−0.1914	0.0501	<0.001
	Weight (g)	0.0059	0.0019	0.002
	Ordinal date	−0.1547	0.0561	0.006
	TL (mm) × ordinal date	0.0005	0.0002	<0.001
	GSI × ordinal date	−0.0914	0.0403	0.023

in 65–75% of the cases of failed prediction, depending on the model. Pairwise comparisons of failed predictions for the four models yielded Sørensen's coefficients between 0.75 and 0.89, indicating a high degree of similarity (i.e., non-random) in the individual fish that each model failed to predict correctly (Table 5).

Table 5
Sørensen's similarity coefficients based on pairwise comparisons of failed predictions by the four models presented in Fig. 4. Coefficients are expressed as a percentage and represent the proportion of failures common to both models (i.e., failed maturity predictions for the same individuals in each model).

Model	1234578	234578	23
1234578	–	–	–
234578	88.7%	–	–
23	75.8%	80.0%	–
2	78.5%	77.4%	87.1%

For a subset of fish, we also investigated the performance of novel predictors—gonad size, gonad color, and capture location. For the top ten AICc-ranked models, all novel predictors with the exception of longitude improved mean classification success by at least 5%. When added to each of the best jackknife reclassification models for a given number of predictor variables, the novel predictors resulted in more modest improvement (about 3%) in classification success, again with the exception of longitude. Similar novel predictors produced comparable model improvements; e.g., both measures of gonad color performed similarly, as did the three measures of gonad morphology. Overall, the addition of one or more novel predictors resulted in several good models (regardless of the ranking criteria), which were able to predict maturity correctly >90% of the time.

3.2. Hindcasting variation in maturity schedules

The simplest model that achieved the best jackknife reclassification success was selected for use in hindcasting southern flounder maturity schedules. Two models achieved >87% classification success (see Table 3 for reclassification performance not including novel predictors) and model 348 (maturity = $\beta_0 + \beta_1 \times \text{TL} + \beta_2 \times \text{weight} + \beta_3 \times [\text{GSI} \times \text{ordinal date}]$) was selected because of the reduced number of predictor terms. It should be noted that model 348 and model 23457 (the two models with >87% classification success) contained nearly identical information. Model 23457 included each of the predictors included in model 348, with the only difference between the models being that model 348 included the interaction between GSI and ordinal date of capture, while model 23457 included each of those predictors separately. Therefore, we expected these two models to generate similar retrospective estimates of southern flounder maturity ogives, and focused our analysis on the simpler model. In

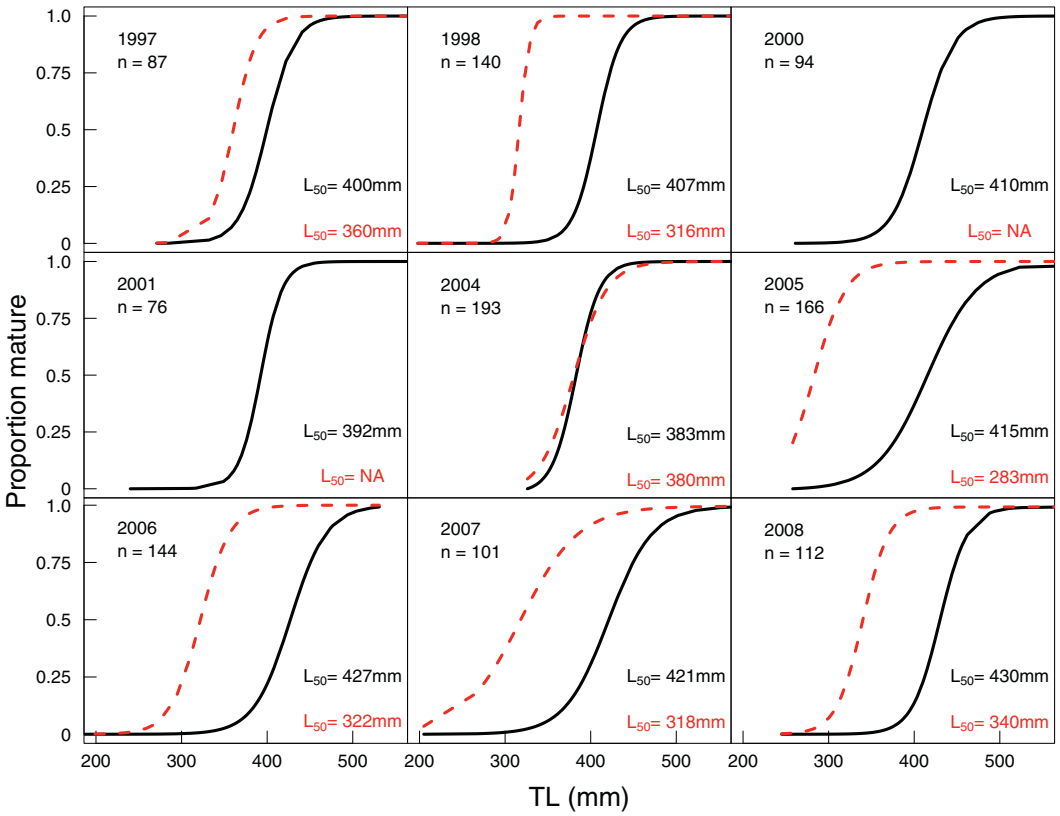

Fig. 5. Estimated maturity ogives for nine years (1997, 1998, 2000, 2001, and 2004–2008) using probabilistic model hindcasting (Model 348; solid black line) and a conventional macroscopic-only model (Model 1; dashed red line). Model 348 included TL, weight, and the GSI × ordinal date interaction, and in each panel the L_{50} estimate for Model 348 is presented in black text above the L_{50} estimate for Model 1. Note: Macroscopic data from 2000 and 2001 were insufficient in contrast to produce maturity ogives, and are not reported.

addition, the simpler model slightly outperformed the more complex model during 2-fold (annual) cross validation (see Table 3). Sufficient data to calculate a maturity ogive were available from nine years: 1997, 1998, 2000, 2001, and 2004–2008, producing a mean L_{50} = 409 mm TL with a range of 47 mm (L_{50} = 383–430 mm TL; Fig. 5). In cases when L_{50} estimates were calculated for consecutive years, the mean difference between any two adjacent years was 14 mm (range = 6–32 mm). When maturity ogives were reconstructed using only traditional macroscopic stages (i.e., using model 1, where 1 = assigned macroscopic stage), the range of L_{50} estimates was twice as large (97 mm) and all were predicted to be ≤380 mm TL, outside of the entire range of L_{50} estimates generated by our best jackknife reclassification model.

3.3. Whole mount potential

Based on recorded observations made when viewing a training set of ten whole mount samples paired with histological stages, we generated whole mount descriptions for the three oocyte stages that we encountered (Table 6; Fig. 6). For 105 independently assessed ovarian tissue samples, agreement between histological assignments and whole mounts was high. Based on standard histological preparation and analysis, the sample included 55 individuals possessing only primary growth oocytes, 32 fish with cortical alveolar oocytes, and 18 with vitellogenic oocytes. Comparatively, whole mount staging identified 51 fish with primary growth oocytes, 37 with cortical alveolar oocytes, and 17 possessing vitellogenic oocytes. Although stage-specific agreement was high, correspondence between the methods improved further when fish were simply classified as mature or immature, which is the primary level of classification required to construct maturity schedules. Using whole mounts, we were able to correctly stage 93% of immature and 100% of mature southern flounder. Although some cortical alveolar and vitellogenic oocytes were confused, the presence of either stage was still considered to indicate a mature fish. Only four individuals were incorrectly assigned mature or immature status by the whole mount approach and each instance involved a fish possessing only primary growth oocytes (immature) based on histology which was incorrectly staged as having cortical alveolar oocytes (mature) using whole mounts.

Table 6
Descriptions of primary growth, cortical alveolar, and vitellogenic oocyte stages when viewed using whole mount methods.

Stage	Description
Primary growth	Oocytes clear; nucleus visible; few/none clear Balbiani bodies randomly distributed; oocytes packed densely and very cohesive
Cortical alveolar	Increase in size over primary growth; small cortical alveoli arranged as dark clusters around nucleus; zona radiata in more developed samples, but not recommended for positive identification
Vitellogenic	Oil droplets larger than cortical alveoli and packed throughout oocyte (to edge); well-defined zona radiata; considerably larger than other stages; generally dense and dark appearance; to the naked eye or with a dissecting microscope oocytes appear yellow and singular (not cohesive as in earlier stages)

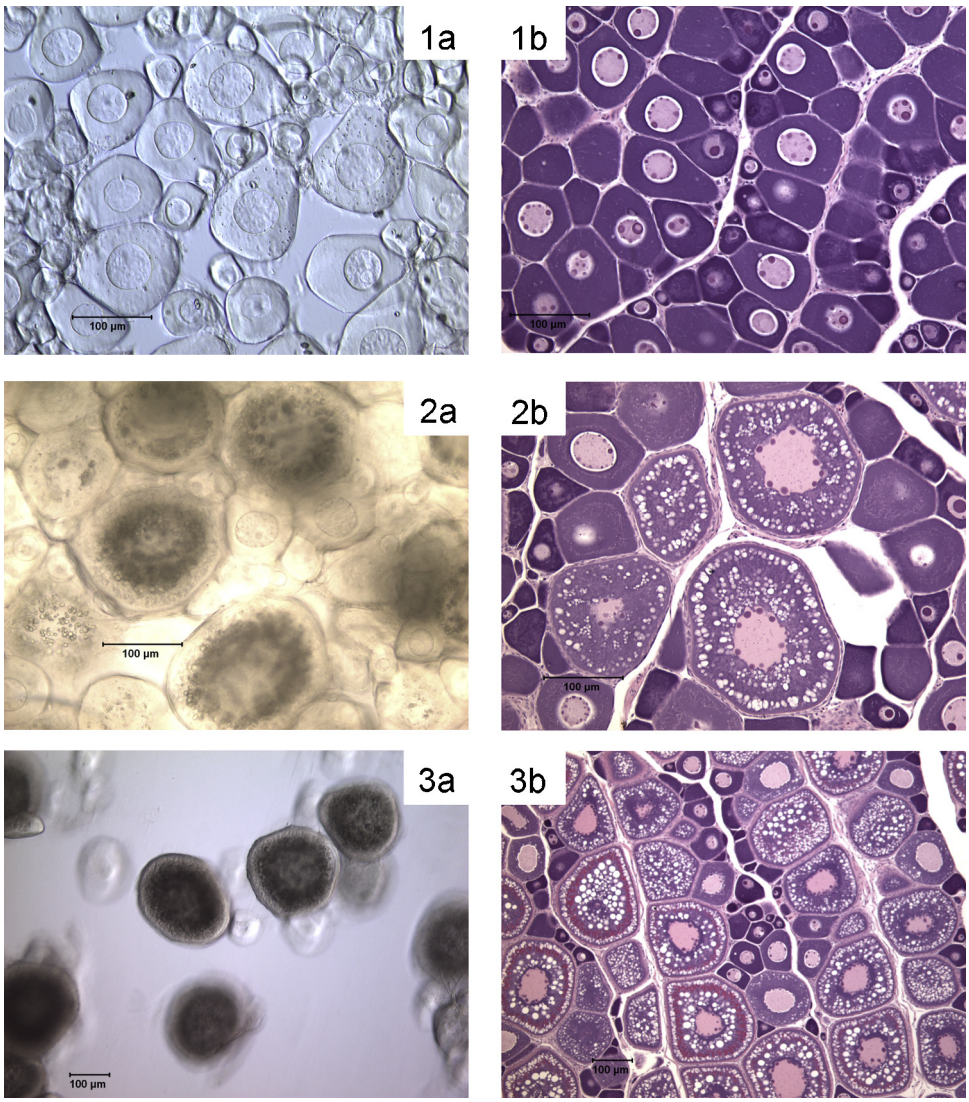

Fig. 6. Examples of whole mount (a) and histological (b) images taken from three southern flounder. Shared numbers indicate the same individual fish. 1a and 1b are each examples of primary growth stage oocytes; 2a and 2b are each examples of cortical alveolus stage oocytes (based on Grier et al., 2009); 3a and 3b are each examples of vitellogenic stage oocytes.

4. Discussion

4.1. Predicting maturity

To our knowledge, adult southern flounder have never been collected from their offshore spawning locations. Assessing the maturity of females is thus restricted to pre-spawning individuals, many of which only possess oocytes in early stages of development. Consequently, recent findings have highlighted potential errors in the estimation of maturity schedules for this species using only macroscopic staging criteria (Midway and Scharf, 2012). In this study we have demonstrated that through predictive modeling considerable improvements in the accuracy of maturity assignments can be achieved. We developed and tested a set of multivariable models capable of predicting maturity for southern flounder females with up to 90% success with minimal additional data collection requirements. A major benefit of the modeling approach we used is that it can be done at very low cost since most of the input data was already available and the models were constructed and cross-validated using free software (R Development Core Team, 2012). Although we did complete histological assessment for each fish to test our models, the resources necessary to complete histology may only be required during model development. After development and testing, models may require periodic re-calibration to adjust for changes in the relationship between model predictors (e.g., age and size) and maturity due, for example, to selective harvest pressure. However, fisheries-induced evolution is unlikely to produce large changes in life history traits within relatively short time periods (Hilborn and Minte-Vera, 2008; Andersen and Brander, 2009), meaning re-calibration should be necessary infrequently. Additional histological analyses may be required for fishes with high error rates, such as those in very early stages of maturity, but this should be limited to small subsets of individuals in most cases. The use of predictive modeling for fish maturity assignment has been limited historically; however, our findings, along with other recent examples (Vitale et al., 2006; Schill et al., 2010; Peer et al., 2012), imply that reliable estimates of maturity can be generated using mostly data that is already being collected. We contend that multivariable predictive modeling of fish maturity holds great promise as an improvement over traditional macroscopic assessment, and its application should be explored for other managed species.

We found that a large number of models achieved high classification success rates, with roughly three-fourths of the models

tested being able to successfully predict maturity for >83% of fish in our dataset. While most of the best models included multiple predictor variables, GSI was included in all of the best AICc and cross-validation models, demonstrating that sufficient contrast in GSI within the dataset can help to generate accurate maturity predictions. Indeed, a model including only GSI as a predictor was able to correctly assign maturity status nearly 85% of the time for southern flounder. Historically, GSI has mainly been used to document reproductive seasonality rather than to assign maturity, however recent investigations have begun to highlight its potential to contribute to maturity schedules. Tomkiewicz et al. (2003) showed that GSI aligned well with revised oocyte stages in the Baltic stock of Atlantic cod (*G. morhua*), while Vitale et al. (2006) were able to use GSI as a main predictor variable in a regression tree model that predicted Kattegat cod maturity with high rates of success. Likewise, McPherson et al. (2011) concluded that GSI could be used to fine tune macroscopic maturity assignments for Atlantic herring (*Clupea harengus*) after detecting underestimation of SSB by as much as 27% when maturity was assigned using macroscopic traits alone. The southern flounder included in our dataset were captured prior to offshore emigration and thus, generally did not exhibit large values of GSI. However, we found that sufficient contrast existed in GSI between immature and mature individuals to enhance our predictive ability, which is consistent with the recent findings implicating GSI as a valuable and easily obtained predictor of fish maturity.

In our models, errors associated with GSI as a predictor occurred mostly for mature fish having intermediate values of GSI, which were classified incorrectly as immature. Similarity indices revealed that model failures occurred mainly for the same individuals (75–89% of failures in common) across different model configurations. The fact that these failures occurred in several models which included multiple predictor variables beside GSI indicates that early developing gonadal stages will likely still require supplemental histological analysis, depending on species. A closer examination of model performance for a macroscopic-only model provided further evidence that southern flounder early in reproductive development will likely require additional analysis. While the model that included only the assigned macroscopic stage as a predictor variable was able to achieve a 79% success rate, this performance should be interpreted cautiously. Specific macroscopic stages (e.g., developing) for southern flounder had very high error rates (up to 39%), meaning that the predictive ability of any model based only on assigned macroscopic stages will be a function of the distribution of macroscopic stages within the modeled dataset. Models built using datasets consisting of mostly fish in early gonadal development stages, which tend to be the most error prone, will generally achieve poorer success rates when the assigned macroscopic stage is the sole predictor variable. Diminished predictive capability is likely to be common among fishes when individuals early in the maturation process contribute most to available data sets.

The novel predictor variables that we generated enabled our best models to achieve >90% prediction success rates. We found that gonad color, a traditional trait used for macroscopic staging, and gonad dimension could each be quantified in simple ways to improve maturity assignments. Peer et al. (2012) recently demonstrated success when linking ovary color to oocyte stages for striped bass, and concluded that quantification of ovary color could likely be used to aid maturity assignments for many species. The growing availability of low cost digital imaging capabilities has led to increased use in fisheries (Cadrin and Friedland, 1999), and should enable more widespread application of these techniques. For southern flounder, both the ovary color analysis and the measurement of gonad dimensions involved one additional step when processing each fish that required minimal time and little to no cost. Additional variables such as these could be obtained more routinely as a way to increase accuracy of fish maturity assignments.

Despite their lack of widespread use in assessing maturity, we found that simple whole mounting of gonadal tissue was very successful when applied to southern flounder, with close to 95% agreement with histological assignments. Our findings were aligned with those from previous studies that also demonstrated high levels of agreement between whole mount techniques and histology (Forberg, 1983; West, 1990). The main limitation we encountered when using whole mount methods regarded the identification of transitional oocyte stages (this was also noted by West, 1990), particularly oocytes transitioning between primary growth and cortical alveolar stages. It should be noted, however, that accurate identification of early oocyte transitions such as this one can be problematic even when examining histological preparations, and that perhaps no method can expected to be 100% accurate. During the transition to secondary growth, small features can often be detected within the oocytes, but they have yet to develop the proteins that would be cross-linked by a formalin preservative (Kiernan, 2000) to give whole mount oocytes the opaqueness that is used to characterize them as developing or secondary growth. Despite this, we found that many early cortical alveolar stage oocytes were identifiable in whole mount preparations for southern flounder, so otherwise problematic transitional stages may be successfully analyzed for this species. Whole mount methods could be used to complement predictive models to improve overall classification success to close to 100%. For instance, fish having intermediate GSI values, which were the most frequently misclassified by the models, could be identified during the modeling process and subject to whole mount methods. The use of whole mount techniques, combined with predictive modeling, may represent an efficient approach for improved maturity assessment and reduce the need for histology.

4.2. Model selection

AIC scores and jackknife reclassification success rates were strongly correlated across the range of models we tested, an indication that several routinely collected biological variables can inform predictions of maturity, especially for fish often captured in early stages of reproductive development like southern flounder. However, the correlation between model assessment tools was weaker when comparisons were restricted to the better models. That is, for models with both AIC scores <350 and jackknife reclassification success rates >83% (which comprised about three-fourths of all models), the model rankings between AIC and jackknife reclassification were not aligned as well. The disparities in model rankings reflect fundamental differences between the model assessment approaches, but also have practical implications that require consideration (Burnham et al., 2011). AIC model selection ranks models based on their relative probabilities given the existing data, elevating models that couple minimal information loss with parsimony (Burnham and Anderson, 2002). Model validation approaches, such as jackknife reclassification, differ fundamentally in that they rank models solely based on their performance in correctly classifying an independent observation (Shao, 1993). In our study, jackknife reclassification performance benefitted from the binary classification of maturity, and also from the fact that a model success occurred if the predicted probability of maturity was >50% or <50% for females identified as mature, or immature, respectively, using histological techniques. Therefore, we were not surprised that several simple models containing relatively few predictors could achieve high reclassification success. However, closer inspection of the deviations between model predictions and actual maturity assignments (based on histology) revealed that most deviations were small and that they were relatively similar in magnitude between simple (few predictors) and complex (many predictors) models. Basically, more complex models did not lessen model

deviations appreciably and did not prevent reclassification errors, relative to simpler models.

For the data sets we modeled, improvements in model fit that were gained by the inclusion of many predictors exceeded the penalties for model complexity, resulting in the lowest AIC scores for more complex models. Although model fit was improved, the data requirements and associated costs for more complex models may not be trivial in some instances. There were models in our set that required fewer (2–4) predictors while still receiving reasonable support (ΔAICc $\leq$ 10), and the data required for these simpler models may be more attainable for a range of exploited species. In addition, since AIC rankings are based solely on model fit to the existing data, the most supported models are often fine tuned to specific attributes of the data used to build the model and may lack generalizability. For predicting fish maturity, AIC model selection approaches may be most appropriate for cases when the predictors are known to be (relatively) invariant over time, which in uncommon for most stocks (Lowerre-Barbieri et al., 2011a). Our findings support the evaluation of multiple model assessment approaches, which vary in their objectives, to identify a set of models that combine model fit with good prediction performance when applied to independent data.

4.3. Retrospective analysis

Temporal changes in maturity schedules are well documented among fishes (e.g., Rijnsdorp, 1993; Morgan and Colbourne, 1999; Walsh and Morgan, 1999; Wang et al., 2008) and have often been attributed to density-dependent changes in growth rates resulting from exploitation. Additionally, variability in life history traits, like size-at-maturity, can be caused by environmental (e.g., temperature), ecological (e.g., prey availability) and genetic variation; however, it is often difficult to distinguish the importance of multiple factors (Rose et al., 2001; Marshall and Browman, 2007). Even with stable population abundance, shifts in demography can lead to detectable variation in fish maturity schedules (Wright and Trippel, 2009). Regardless of the cause of variation in the timing of maturity, its impact on the response of a stock to harvest can be considerable and population assessments would benefit from a better understanding of its magnitude.

The model that we used to hindcast southern flounder maturity schedules did not include the assigned macroscopic stage as a predictor variable, meaning it was free from any maturity assignment errors that may have occurred when fish were originally examined. The mean L_{50} estimate for the nine years analyzed from the historical dataset (409 mm TL) was nearly identical to the estimate of L_{50} (408 mm TL) generated from a recent comprehensive histological analysis (Midway and Scharf, 2012). The close agreement between the historical and contemporary estimates of the average size at maturity provides some measure of confidence in the model and its ability to quantify temporal variability in southern flounder maturity schedules. We detected a range in L_{50} estimates of 47 mm among nine years, with adjacent years generally closer in their estimates. Interannual variation in L_{50} by as much as 1–2 inches can impact yearly estimates of SSB and thus, target harvest rates for fisheries whose management includes minimum size limits, such as the southern flounder fishery in NC. Our retrospective examination of southern flounder maturity schedules also illustrated the bias of traditional macroscopic maturity assignment toward classifying smaller fish as mature. When compared to the hindcast model (model 348), a model using only the assigned macroscopic stage as a predictor variable (model 1) produced much lower L_{50} estimates (all years $\leq$380 TL) and a range that was nearly twice as broad by comparison, an indication of the likely sensitivity of the model to the distribution of macroscopic stages within each annual data set.

5. Conclusions

There has been increased recognition of the importance of reproductive information for fishery population assessment (Kjesbu, 2009; Lowerre-Barbieri et al., 2011b). New information is challenging long-held assumptions, such as relatively constant maturity schedules (Rochet, 2009). Integrating more and better reproductive information into stock assessments is becoming a goal of many management agencies, yet time and expense often preclude the customary use of methods such as histology. Multivariable predictive models that take advantage of traditionally collected biological information may provide fishery managers with a tool to generate more accurate maturity assignments with minimal investment. Additionally, oocyte whole mounting may prove to be a fast, inexpensive, and reliable method to validate the maturity predictions from models and may reduce the need for investing in histological analyses (West, 1990; Neidig et al., 2000; Lowerre-Barbieri et al., 2011b). Lastly, predictive models that perform well during cross validation can be a promising approach to conduct retrospective analysis of variability in maturity timing, which can enable contemporary fluctuations to be interpreted within a broad context.

Acknowledgements

We thank commercial fishers Troy Outland, Cecil Simons, and Jeff Wolfe for assistance with fish collections. Staff members Chip Collier, Lee Paramore, Katy West, McLean Seward, Lele Judy, and Jason Rock generously donated time to collect samples for us as part of the NCDMF fishery-independent sampling program. Jon Vanderfleet, Megan Davis, and Nikolai Klibansky assisted with fish processing. Two anonymous reviewers provided helpful suggestions that improved the presentation of our ideas. Funding for this research was provided by North Carolina Sea Grant through the Fishery Resource Grant Program. Interpretation of these data and any conclusions drawn are those of the authors and do not necessarily represent the views of the North Carolina Division of Marine Fisheries.

References

Andersen, K.H., Brander, K., 2009. Expected rate of fisheries-induced evolution is slow. Proc. Natl. Acad. Sci. U.S.A. 106, 11657–111660.
Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed. Springer-Verlag, New York.
Burnham, K.P., Anderson, D.R., Huyvaert, K.P., 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav. Ecol. Sociobiol. 65, 23–35.
Brown-Peterson, N.J., Wyanski, D.M., Saborido-Rey, F., Macewicz, B.J., Lowerre-Barbieri, S.K., 2011. A standardized terminology for describing reproductive development in fishes. Mar. Coast. Fish.: Dynam. Manage. Ecosyst. Sci. 3, 52–70.
Cadrin, S.X., Friedland, K.D., 1999. The utility of image processing techniques for morphometric analysis and stock identification. Fish. Res. 43, 129–139.
Costa, A.M., 2009. Macroscopic vs. microscopic identification of the maturity stages of female horse mackerel. ICES J. Mar. Sci. 66, 509–516.
Dhillon, R.S., Fox, F.G., 2004. Growth-independent effects of temperature on age and size at maturity in Japanese medaka (*Oryzias latipes*). Copeia 2004, 37–45.
Ferreri, R., Basilone, G., D'Elia, M., Traina, A., Saborido-Rey, F., Mazzola, S., 2009. Validation of macroscopic maturity stages according to microscopic histological examination for European anchovy. Mar. Ecol. 30, 181–187.
Forberg, K.G., 1983. Maturity classification and growth of capelin, *Mallotus villosus villosus* (M), oocytes. J. Fish Biol. 22, 485–496.
Grier, H.J., Uribe-Aranzábal, M.C., Patiño, R., 2009. The ovary, folliculogenesis and oogenesis in teleosts. In: Jamieson, B.G.M. (Ed.), Reproductive biology and phylogeny of fishes (agnathans and bony fishes), Volume 8A. Science Publishers, Enfield, New Hampshire, pp. 25–84.
Hilborn, R., Minte-Vera, C.V., 2008. Fisheries-induced changes in growth rates in marine fisheries: are they significant? Bull. Mar. Sci. 83, 95–105.
Hunter, J.R., Macewicz, B.J., 1985. Rates of atresia in the ovary of captive and wild northern anchovy, *Engraulis mordax*. Fish. Bull. 83, 119–136.
Hutchings, J.A., Baum, J.K., 2005. Measuring marine fish biodiversity: temporal changes in abundance, life history, and demography. Philos. Trans. R. Soc. 360, 315–338.

Kiernan, J.A., 2000. Formaldehyde, formalin, paraformaldehyde and glutaraldehyde: what they are and what they do. Microsc. Today 00-1, 8–12.
Kjesbu, O.S., 1991. A simple method for determining the maturity stages of Northeast Arctic Cod (*Gadus morhua*) by in vitro examination of oocytes. Sarsia 75, 335–338.
Kjesbu, O.S., 2009. Applied fish reproductive biology: contribution of individual reproductive potential to recruitment and fisheries management. In: Jakobsen, T., Fogarty, M.J., Megrey, B.A., Moksness, E. (Eds.), Fish Reproductive Biology: Implications for Assessment and Management. Wiley–Blackwell, Chichester, UK, pp. 293–332.
Law, R., Grey, D.R., 1989. Evolution of yields from populations with age-specific cropping. Evol. Ecol. 3, 343–359.
Lowerre-Barbieri, S.K., Barbieri, L.R., 1993. A new method of oocyte separation and preservation for fish reproduction studies. Fish. Bull. 91, 165–170.
Lowerre-Barbieri, S.K., Ganias, K., Saborido-Rey, F., Murua, H., Hunter, J.R., 2011a. Reproductive timing in marine fishes: variability, temporal scales, and methods. Mar. Coast. Fish. 3, 71–91.
Lowerre-Barbieri, S.K., Brown-Peterson, N.J., Murua, H., Tomkiewicz, J., Wyanski, D.M., Saborido-Rey, F., 2011b. Emerging issues and methodological advances in fisheries reproductive biology. Mar. Coast. Fish. 3, 32–51.
Marshall, C.T., Browman, H.I., 2007. Disentangling the causes of maturation trends in exploited fish populations. Mar. Ecol. Prog. Ser. 335, 249–251.
McPherson, L.R., Ganias, K., Marshall, C.T., 2011. Inaccuracies in routinely collected Atlantic herring (*Clupea harengus*) maturity data and correction using a gonadosomatic index model. J. Mar. Biol. Assoc. U.K. 91, 1477–1487.
Midway, S.R., Scharf, F.S., 2012. Histological analysis reveals larger size-at-maturity for southern flounder (*Paralichthys lethostigma*) with implications for biological reference points. Mar. Coast. Fish.: Dynam. Manage. Ecosyst. Sci. 4, 628–638.
Morgan, M.J., 2008. Integrating reproductive biology into scientific advice for fisheries management. J. Northw. Atl. Fish. Sci. 41, 37–51.
Morgan, M.J., Colbourne, E.B., 1999. Variation in maturity-at-age and size in three populations of American plaice. ICES J. Mar. Sci 56, 673–688.
Murua, H., Saborido-Rey, F., 2003. Female reproductive strategies of commercially important fish species in the North Atlantic. J. Northw. Atl. Fish Sci. 33, 23–32.
Neidig, C.L., Skapura, D.P., Grier, H.J., Dennis, C.W., 2000. Techniques for spawning common snook: broodstock handling, oocyte staging, and egg quality. N. Am. J. Aquacult. 62, 103–113.
Peer, A.C., Selckmann, G.M., Miller, T.J., 2012. A standardized method and analytical approach for predicting female reproductive stage in teleosts by using ovary color and female characteristics. Trans. Am. Fish. Soc. 141, 1036–1044.
R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, http://www.R-project.org
Rijnsdorp, A.D., 1993. Fisheries as a large-scale experiment on life-history evolution: disentangling phenotypic and genetic effects in changes in maturation and reproduction of North Sea plaice, *Pleuronectes platessa* L. Oecología 96, 391–401.
Rochet, M.J., 2009. Effects of fishing on the population. In: Jakobsen, T., Fogarty, M.J., Megrey, B.A., Moksness, E. (Eds.), Fish Reproductive Biology: Implications for Assessment and Management. Wiley–Blackwell, Chichester, UK, pp. 172–204.
Rose, K.A., Cowan, J.H., Winemiller, K.O., Myers, R.A., Hilborn, R., 2001. Compensatory density dependence in fish populations: importance, controversy, understanding and prognosis. Fish Fish. 2, 293–327.
Schill, D.J., LaBar, G.W., Mamer, E.R.J.M., Meyer, K.A., 2010. Sex ratio, fecundity, and models predicting length at sexual maturity of redband trout in Idaho desert streams. N. Am. J. Fish. Manage. 30, 1352–1363.
Shao, J., 1993. Linear model selection by cross-validation. J. Am. Stat. Assoc. 88, 486–494.
Sørensen, T., 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. Biol. Skrift./Kongel. Danske Videnskab. Selskab 5, 1–34.
Symonds, M.R.E., Moussalli, A., 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behav. Ecol. Sociobiol. 65, 13–21.
Takade-Heumaker, H., Batsavage, C., 2009. Stock Status of North Carolina Southern Flounder (*Paralichthys lethostigma*). North Carolina Department of Environment and Natural Resources, Division of Marine Fisheries, Morehead City, NC http://portal.ncdenr.org/web/mf/fmps-under-development
Thorsen, A., Kjesbu, O.S., 2001. A rapid method for the estimation of oocyte size and potential fecundity in Atlantic cod using computer-aided particle analysis system. J. Sea Res. 46, 295–308.
Tobin, D., Wright, P.J., 2011. Temperature effects on female maturation in a temperate marine fish. J. Exp. Mar. Biol. Ecol. 403, 9–13.
Tomkiewicz, J., Tybjerg, L., Jespersen, Å., 2003. Micro- and macroscopic characteristics to stage gonadal maturation of female Baltic cod. J. Fish Biol. 62 (2), 253–275.
Vitale, F., Svedäng, H., Cardinale, M., 2006. Histological analysis invalidates macroscopically determined maturity ogives of the Kattegat cod (*Gadus morhua*) and suggests new proxies for estimating maturity status of individual fish. ICES J. Mar. Sci 63, 485–492.
Walsh, S.J., Morgan, M.J., 1999. Variation in maturation of yellowtail flounder (*Pleuronectes ferruginea*) on the Grand Bank. J. Northw. Atl. Fish Sci. 25, 47–59.
Wang, H.-Y., Höök, T.O., Ebener, M.P., Mohr, L.C., Schneeberger, P.J., 2008. Spatial and temporal variation of maturation schedules of lake whitefish in the Great Lakes. Can. J. Fish. Aquat. Sci. 65, 2157–2169.
Wenner, C.A., Roumillat, W.A., Moran Jr., J.E., Maddox, M.B., Daniel III, L.B., Smith, J.B., 1990. Investigations on the Life History and Population Dynamics of Marine Recreational Fishes in SC: Part 1. S.C. Dept. Nat. Res., Charleston, SC.
West, G.J., 1990. Methods of assessing ovarian development in fishes: a review. Aust. J. Mar. Freshw. Res. 41, 199–222.
Witthames, P.R., Thorsen, A., Murua, H., Saborido-Rey, F., Greenwood, L.N., Domingues, R.D., Korta, M., Kjesbu, O.S., 2009. Advances in methods for determining fecundity: application of the new methods to some marine fishes. Fish. Bull. 107, 148–164.
Wright, P.J., Trippel, E.A., 2009. Fishery-induced demographic changes in the timing of spawning: consequences for reproductive success. Fish Fish. 10, 283–304.