Title: Improvements to the Stephens-MacCall approach for calculating CPUE from multispecies fisheries logbook data

Author: Kyle Dettloff[1]

[1]National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southeast Fisheries Science Center, 75 Virginia Beach Dr., Miami, FL 33149, USA

Keywords: Index of abundance; CPUE; Logistic regression; Fisheries dependent data; Red snapper

Abstract

Since its publication in 2004, the Stephens-MacCall method has been widely used as an objective approach to select a subset of fisher-reported trip catch and effort data relevant to a given analysis. This paper demonstrates the undesired effects of using a classification cutoff threshold as originally proposed and offers a weighting alternative which makes use of complete information in the dataset and fitted model. Simple spatiotemporal additions to the original model resulting in improved predictive performance are also presented. These modifications are illustrated with application to commercial red snapper reported in the U.S. South Atlantic and Gulf of Mexico.

1. Introduction

Generating unbiased indices of abundance using fishery-dependent data requires a mechanism of identifying trips that are likely to have targeted a species of interest. That is, due to the positives-only nature of fisher reports, a mechanism must be employed to classify trips on which a species was likely targeted but not caught or caught but not necessarily targeted. To address this difficulty, Stephens and MacCall (2004) proposed a method to estimate the probability that a given trip occurred in the habitat of a target species based on the total species composition of the trip using a Bernoulli generalized linear model (GLM). Only those trips classified as targeted according to some threshold are retained for calculating catch-per-unit-effort (CPUE). This procedure has been widely used as a trip selection approach in the southeastern United States. It was first used for the 2006 stock assessments of Gulf of Mexico vermilion snapper, greater amberjack, and gray triggerfish (SEDAR 2006), and is still employed in current stock assessments (SEDAR 2020a, b) and other research (Carruthers et al., 2015; Ducharme-Barth et al., 2018).

While simple and effective, the subjectivity of selecting a threshold has been identified as a primary drawback to this approach (Thorson et al., 2016). As an alternative to deciding on an arbitrary cutoff, a more defensible replacement for subsetting trips is proposed here which makes use of the complete set of model estimated target probabilities as weights on the full dataset directly in the calculation of CPUE. In addition, the Stephens-MacCall model was originally developed for use in a recreational fishery with unknown fishing location. Accounting for spatiotemporal variation has been shown to reduce bias in abundance indices (Thorson et al. 2016). Given that information on trip date, fishing area, and fishing depth is collected on Southeast Coastal Logbook forms, the positive predictive effect of these variables as additional covariates is illustrated with respect to target probabilities. The impact of these modifications on estimated CPUE is illustrated for red snapper as reported by federally permitted vessels in the southeastern U.S. Atlantic (hereafter Atlantic) and Gulf of Mexico in 2019.

While more sophisticated techniques have been since proposed in the literature to address the issue of CPUE calculation in multispecies fisheries (Winker et al., 2013; Thorson et al., 2015), the suggested modifications to the Stephens-MacCall classification threshold approach as well as some simple spatiotemporal additions to improve model performance will have a direct impact on current and future fish stock assessments in the southeastern United States and other regions due to the method's ubiquity.

2. Materials and methods

2.1 Original model

Stephens and MacCall (2004) proposed a logistic regression on binary presence-absence of a given species of interest y_j with presence-absence of all other potentially co-reported species (1, 2, …, k) as binary predictors to estimate the probability $\hat{p}$ that trip j occurred in the habitat of the target species (Eq. 1).

$$\hat{p}_j = \frac{1}{1 + e^{-\sum_{i=0}^{k} \beta_i x_{ij}}} \tag{1}$$

A classification threshold p_{crit} is then determined, at which trips with higher model predicted probabilities are defined as "targeted" and those with lower model predicted probabilities are defined as "non-targeted". The original text proposes an empirical determination of the classification threshold p_{crit} based entirely on the in-sample misclassification rate of the fitted model. That is, it intends to minimize the sum of the false positive and false negative predictions based on predicted target species presence-absence $\hat{y}_j$ given p_{crit}, with the observed target species presence-absence y_j on each trip defined as the truth (Eq. 2).

$$argmin_{p_{crit}}\left(\sum_j \left| y_j - (\hat{y}_j \mid p_{crit}) \right|\right) \tag{2}$$

There is a discrepancy, however, in the way this calculation is carried out, as the described procedure actually minimizes the difference between total true-positive trips and total predicted positive trips (Eq. 3), rather than minimizing the total misclassification rate.

$$argmin_{p_{crit}}\left(\left| \sum_j y_j - \sum_j \hat{y}_j \mid p_{crit} \right|\right) \tag{3}$$

Differences in threshold behavior are visualized by applying each method to artificial Bernoulli data with a single continuous predictor constructed to provide a simple 2D illustration of how each classification scheme can lead to different results, while also showing how each can be derived independently of the fitted logistic model (Fig. 1). The original procedure was applied to the 2019 commercial red snapper data to compare performance with revised models (Tbl. 1). Model fitting was conducted using the *glm* function with family = "binomial" in R 4.0.4 (R Core Team, 2021).

2.2 Model fitting improvements

Prior to assessing the impact of classification alternatives, a variety of modifications to the original model were considered to potentially improve performance for estimating predicted probabilities. Considered changes included the use of species weight (both square root transformed and relative proportions) instead of simply presence as predictor variables in the regression to see if catch magnitudes offered improvements over simple presence-absence, the addition of two-way interactions to capture how species potentially interact to associate with target probabilities, the addition of a vessel random effect to account for the grouped nature of trips taken on the same vessel, and the addition of spatiotemporal terms (fishing region, month, fishing depth and two-way interactions) to account for how target probabilities differ with space and time.

The efficacy of these changes were assessed as applied to a variety of snapper-grouper species reported in 2019 NOAA Southeast Coastal Fisheries Trip Report Forms (Coastal Logbook, OMB: 0648-0016). Data consist of trip level information including vessel identifier and trip start and unload dates, along with fishing area (allowing for classification by region) and fishing depth for each species. The 2019 version of the coastal logbook form contains 62 listed species with space for write-ins. This allows models to be built with a target species of interest as the binary response while using all remaining finfish species as predictors.

A final set of models with the successful modifications are presented for combined Atlantic and Gulf of Mexico red snapper and compared by Akaike's Information Criterion (AIC). While each region (Atlantic, Gulf of Mexico) is traditionally modeled separately in Southeast red snapper stock assessments for various biological and management reasons, the regions were kept combined here as an illustrative exercise to allow the explicit demonstration of regional differences through interaction terms. For simplicity, gear type was also not considered due to complexities that arise when dealing with multiple gear types on the same trip. Predictive performance and discriminative ability of the final selected model was compared to that using the original Stephens-MacCall methodology as assessed by McFadden's pseudo R^2 and Somers' D, respectively (Tbl. 1).

2.3 Classification alternatives

To assess the performance of the logistic regression as a classification model, the proportion of predicted probabilities that fell very close to zero or one was examined. A large proportion of intermediate probabilities would be an indication the model does not have enough information to make confident binary classifications. Trips were assigned to one of three categories of prediction confidence (non-targeted, possibly targeted, very likely targeted) based on the points of diminishing return on the empirical cumulative distribution function (eCDF) (Fig. 2) of the model predicted probabilities, defined here as

$$argmax_x(F(x) - x) \tag{4}$$

for the lower bound, and

$$argmin_x(F(x) - x) \tag{5}$$

for the upper bound.

Trips with predicted probabilities falling at or below the lower bound were determined to be confident negatives (non-targeted), trips at or above the upper bound confident positives (very likely targeted), and trips in between uncertain (possibly targeted). The total number of trips falling in each category is presented for both the original and revised red snapper models (Tbl. 2).

Rather than rely on a binary classification rule where all trips above a threshold are treated as certainly targeted and all those below are excluded entirely, a natural solution that incorporates target uncertainty among all trips is to use the model estimated probabilities directly as precision weights for each observation when calculating an index. This has the effect of giving full weight to trips that are very likely targeting a species ($\hat{p}$ close to 1) while effectively excluding trips with $\hat{p}$ close to zero from the calculation. All trips are retained with influence proportional to their estimated target probabilities. In a regression setting, this amounts to minimizing the weighted sum of squares when calculating parameter values (Eq. 6) and associated variances (Eq. 7):

$$\hat{\beta} = (\boldsymbol{x}^T\boldsymbol{w}\boldsymbol{x})^{-1}\boldsymbol{x}^T\boldsymbol{w}\boldsymbol{y} \quad (6)$$

$$var(\hat{\beta}) = \hat{\sigma}^2(\boldsymbol{x}^T\boldsymbol{w}\boldsymbol{x})^{-1} \quad (7)$$

where *x* is a vector of the explanatory variable (e.g., effort), *y* is the response vector (e.g., catch), *w* is a vector of the weights (estimated target probabilities).

The estimated residual variance for a weighted regression with *p* parameters (Eq. 8) is defined as:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} w_i(y_i - \hat{y}_i)^2}{n-p} \quad (8)$$

In the simplest case of an intercept only model, $\hat{\beta}$ is equivalent to calculating a weighted arithmetic mean of the observed values (Eq. 9):

$$\bar{y} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \quad (9)$$

This approach was applied to the 2019 red snapper presence-absence data using the *lm* function in R to fit an intercept only model (i.e., assuming all trips had equal effort), with the logistic model predicted probabilities supplied to the “weights” argument. The same principle could be employed in a GLM setting (*glm* function), where the supplied weights indicate the inverse dispersion of each observation. In a practical setting, the presence-absence response variable would be replaced with total catch and the intercept with an effort measure.

Note that estimated weights below a set tolerance may need to be manually set to zero to ensure that very small decimals arising from machine floating point precision are not included in degrees of freedom calculations when variance is of concern. Calculating the weighted estimate using only trips with estimated probabilities falling above the lower bound in Eq. (4) could be a practical way to achieve this with minimal impact on estimates, depending on the distribution of predicted probabilities. The impact of this on estimated CPUE for red snapper is presented and contrasted against estimates obtained from subsetting the dataset according to various binary thresholds (Tbl. 3).

3. Results

3.1 Model fitting improvements

The following considered changes were abandoned either due to lack of improvement or infeasible implementation in test scenarios.

1. *Use of (square root transformed) species weights instead of presence-absences as predictors*:

   This resulted in reduced predictive performance according to AIC. The additional noise introduced by using total weight or square root transformed outweighed the benefit of the added information.
2. *Use of proportions of total co-occurring species weight instead of presence-absence as predictors*:

   This also resulted in no improvement in predictive performance, and often models did not converge.
3. *Addition of (2-way) interaction terms to presence-absence main effects*:

   Interactions proved computationally burdensome to implement. For example, a logbook with 175 possible non-target species produces $\binom{175}{2}$ = 15,225 additional terms to estimate.
4. *Addition of a vessel random effect*:

   Models did not converge successfully.

In contrast, the changes below resulted in models with improved predictive performance and successful convergence under a variety of test scenarios.

1. *The addition of landing month as a categorical predictor*:

   The addition of a month variable resulted in substantially improved models according to AIC. The addition of temporal information at this level resulted in better predictive performance than collapsing to the season level (i.e., winter, spring, summer, fall).
2. *The addition of fishing region as a categorical predictor*:

   The interaction of fishing region with month was included, allowing the effect of landing month to vary by region, for which there was evidence. This may also be a useful consideration for finer-scale fishing areas if this information is available.
3. *The addition of fishing depth as a polynomial spline covariate*:

   Fishing depth can be a valuable proxy of the species being targeted, even if reported depths are not completely accurate. Average fishing depth reported among the non-target species was taken as a trip level covariate, which proved preferable to using the trip level median or maximum. Due to the non-linear relationship of target probability with depth, predictive performance was maximized by including depth as a polynomial spline term (*df* = 25) (Tbl. 1). A depth and region interaction term was selected, allowing the effect of depth on target probability to vary between the Atlantic and Gulf of Mexico. The addition of a depth by month interaction was considered but ultimately excluded due to unsuccessful model convergence.

In quantifying these successful additions, logistic model predictive performance as assessed by AIC for 2019 Southeast coastal trips with red snapper as the species of interest (n = 33,671 region trips; data as of 27 Feb 2021) indicated that the addition of region, month, depth, depth x region and month x region resulted in the best model among those assessed (Tbl. 1). AIC was substantially reduced from the original model relying on only species presence-absence as predictors. Notice the improved separation of the estimated target probabilities toward zero and one in the empirical CDFs under the selected revised model as compared to the original (Fig. 2). While results are only presented for red snapper, it should be noted that similar improvements in model fit were realized for other species (e.g., vermilion snapper, yellowtail snapper, etc.) when the additional spatiotemporal covariates were included.

A comparison between trips selected by the final model in Tbl. 1 (p_{crit} = 0.5) vs. trips selected by the original Stephens-MacCall method (also using p_{crit} = 0.5) shows that the majority of selected trips were common to both models, but among the records unique to each method, the revised model selected a much higher proportion of positive trips, and therefore had an overall higher percent positive rate (85.7% vs. 77.5%). 6,227 trips reported red snapper catch in the observed data. The revised model also exhibited a substantially higher pseudo-R^2 (0.73 vs. 0.50) and Somers' D (0.96 vs. 0.86), suggesting a superior fit and discriminative ability with the added covariates (both range from 0 to 1) (Tbl. 4).

Using p_{crit} as originally calculated (0.405) with the original model compared to the same model with p_{crit} set to 0.5, 1,660 trips predicted to be more likely not targeting than targeting red snapper would have been included in the selected subset (78.6% of which were subsequently selected by the model with additional covariates at p_{crit} = 0.5).

3.2 Classification alternatives

The two classification methods in Eq. (2) and Eq. (3) may lead to different results, as illustrated with artificial data (Fig. 2). It is not clear that anything meaningful is being optimized by Eq. (3); the justification for minimizing the difference between observed positive and predicted positive trips is ambiguous. Observe in Fig. 1 that the total misclassification threshold as defined by Eq. (2) shifts left to capture one additional negative trip to balance the positive trip outlier, and that both thresholds defined by Eqs. (2) and (3) can be identified visually from the observed data, independent of the fitted logistic model.

As is evident in Fig. 2, a majority of total trips under the revised model could be confidently classified as "non-targeted", of which only 0.2% and 1.9% were observed to have caught red snapper as classified by the revised and original models, respectively (Tbl. 2). A much smaller number of trips were classified as "very likely targeted", though a greater number achieved this designation in the revised model due to its superior predictive performance. In both models, the number of trips in the "very likely" category is much lower than the total number of trips with observed red snapper catch (6,227), suggesting this would be an unsuitable cutoff and likely overestimate CPUE. It is clear that trips in the "possibly targeted" category must contribute to the calculation in some way, and rather than being subset according to an arbitrary cutoff, this can be done through trip weighting.

If the full unweighted dataset were used to calculate CPUE (defined here for illustrative purposes as number of trips with red snapper catch over number of total trips), the result is of course biased low (0.185, Tbl. 3), hence the initial need for the Stephens-MacCall approach to identify which trips are likely actually targeting the species of interest. When the calculation is done with the confidently "non-targeted" trips removed (i.e., those with predicted probabilities below the lower point of diminishing returns on the empirical CDF), the estimates become more realistic (original model: 0.504, revised model: 0.547), but are likely still underestimates since full and equal weight is given to every trip in the remaining subset regardless of its predicted target probability.

Likewise, performing the calculation using only trips with predicted probabilities greater than 0.5 lessens the impact of treating all lower probability trips equally (original model: 0.775, revised model: 0.857), but at the same time has the disadvantage of completely discarding all low probability trips, some of which may have been truly targeted. The weighted estimates mitigate these concerns by allowing every trip to influence the calculation directly in proportion to its predicted target probability (original model: 0.622, revised model: 0.787). The higher revised model estimate is reflective of the improved ability of the model to discern between very low target probability (estimates closer to zero) and very high target probability (estimates closer to one) trips, illustrating the importance of selecting an accurate underlying model to generate the predicted target probabilities. Completely removing "non-targeted" trips before weighting had minimal impact on the weighted estimate under the revised model (all trips included: 0.787, non-targeted trips removed: 0.791) (Tbl. 3).

4. Discussion

The original Stephens-MacCall model provides a simple and useful approach for assigning target probabilities to multispecies logbook reported trips based strictly on the presence and absence of potentially co-reported species. Regardless of the inconsistency in critical value calculation, the threshold adjustment approach based on minimizing the in-sample misclassification rate is fundamentally flawed, and amounts to an extreme overfitting to the data which will lead to an incorrect model (Harrell 2015). In fact, decision rules to arrive at the same results as those provided by Eqs. (2) and (3) can be derived simply from the data without even fitting a logistic model. The originally proposed in-sample post-hoc classification threshold should not be used to avoid overriding the logistic model and maximum likelihood estimation entirely, and the cutoff value p_{crit} should either 1) be set to 0.5 as a starting point or determined according to a cost function based on prior expert opinion if a threshold must be used, or 2) be abandoned entirely in favor of using the estimated target probabilities as weights on the complete dataset when calculating CPUE. The second option is recommended as it eliminates the need to select a cutoff which may alter indices substantially depending on how it is defined, but instead makes use of the full dataset without unnecessarily discarding information or forcing the analyst to make a subjective decision. This also eliminates the undesirable effect of applying arbitrarily inconsistent cutoffs among species which may be sensitive to the distribution of modeled probabilities. The weighting of individual observations is widely supported within regression functions across standard statistical software and should prove easy to implement in practice.

Note that the proposed modification to the probability threshold alone is not expected to change indices consistently in any one direction as it can be essentially arbitrary where the Stephens-MacCall p_{crit} value happens to fall relative to 0.5. The three category classification scheme provided further demonstration that estimated CPUE can be very sensitive to which threshold is chosen. These points become irrelevant under the trip weighting alternative. In any scenario, the target probabilities estimated by the model are assumed to be correct. If the underlying model is flawed, estimates of CPUE based on this approach may be biased.

Regarding revisions to data inputs, the use of presence-absence as binary predictors proved preferable to alternate formulations considered such as total species weights or proportions of total weight. Additional model complexity such as vessel random effects and species interactions also proved unsuccessful to implement.

Depth, region, and month, along with the interactions of month by region and depth by region each provided substantial improvement to model predictive performance and resulted in a final model with a higher proportion of positive records selected. This is intuitive, as a particular species is likely to be targeted differently over the course of a year, which itself is likely to differ by region and depth due to both biological and management considerations. The inclusion of these predictors may be used as a starting point for future modeling efforts with data from a broad spatial or temporal range, adding or dropping terms as appropriate for a given scenario to better capture fisher behavior. Evidence for the interaction of region with month and fishing depth also lends additional support to modeling each region separately in stock assessments.

While the interaction of landing month and region may serve as a crude proxy to capture in-season regulations (e.g., closures), temporal regulations should generally be considered explicitly at the appropriate geographic level when available in choosing a candidate frame of data to subset. Additionally, though not considered here, model performance could likely be enhanced by explicitly defining closed periods for predictor species if this information is available; that is, defining each observation according to three possible categories of “closed”, “open but absent”, and “present” rather than simply present/absent, as the closure of a certain group of species may provide valuable information about the target probability of another.

These simple revisions can immediately impact CPUE calculations as applied to multispecies logbook data, resulting in more statistically valid indices of abundance, especially in Southeastern U.S. stock assessments where the method is widely used. Implementation of these changes should also be explored for other regions, species, and data sources that rely on the Stephens-MacCall approach for analyzing multispecies data.

Acknowledgments

I would like to thank Kevin J. McCarthy for helpful input on an early draft. Adyan B. Rios, John F. Walter, Steven G. Smith, and two anonymous reviewers provided constructive feedback and improvements to the manuscript. Thanks are also due to the Miami Coastal Logbook Program staff and database developers for processing and making these data accessible.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Carruthers, T.R., Walter, J.F., McAllister, M.K., Bryan, M.D., 2015. Modelling age-dependent movement: an application to red and gag groupers in the Gulf of Mexico. Can. J. Fish. Aquat. Sci. 72(8), 1159-1176. https://doi.org/10.1139/cjfas-2014-0471

Coastal Logbook: Vessel Trip Report. SE Coastal Fisheries Trip Report Form. OMB Control No.: 0648-0016. https://omb.report/icr/201908-0648-007/doc/94282601

Ducharme-Barth, N.D., Shertzer, K.W., Ahrens, R.N., 2018. Indices of abundance in the Gulf of Mexico reef fish complex: A comparative approach using spatial data from vessel monitoring systems. Fish. Res. 198, 1-13. https://doi.org/10.1016/j.fishres.2017.10.020

Harrell Jr, F.E., 2015. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer.

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

SEDAR, 2006. Stock assessment of the Vermilion Snapper, Greater Amberjack, and Gray Triggerfish in the Gulf of Mexico (SEDAR 09). Charleston, SC.

SEDAR, 2020a. Stock assessment of the Yellowtail Snapper in the Southeastern U.S. (SEDAR 64). North Charleston, SC.

SEDAR, 2020b. Stock assessment of the Vermilion Snapper in the Gulf of Mexico (SEDAR 67). North Charleston, SC.

Stephens, A., MacCall, A., 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fish. Res. 70, 299-310. https://doi.org/10.1016/j.fishres.2004.08.009

Thorson, J.T., Shelton, A.O., Ward, E.J., Skaug, H.J., 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES J. Mar. Sci. 72(5), 1297-1310. https://doi.org/10.1093/icesjms/fsu243

Thorson, J.T., Fonner, R., Haltuch, M.A., Ono, K., Winker, H., 2016. Accounting for spatiotemporal variation and fisher targeting when estimating abundance from multispecies fishery data. Can. J. Fish. Aquat. Sci. 74(11), 1794-1807. https://doi.org/10.1139/cjfas-2015-0598

Winker, H., Kerwath, S.E., Attwood, C.G., 2013. Comparison of two approaches to standardize catch-per-unit-effort for targeting behaviour in a multispecies hand-line fishery. Fish. Res. 139, 118-131. https://doi.org/10.1016/j.fishres.2012.10.014

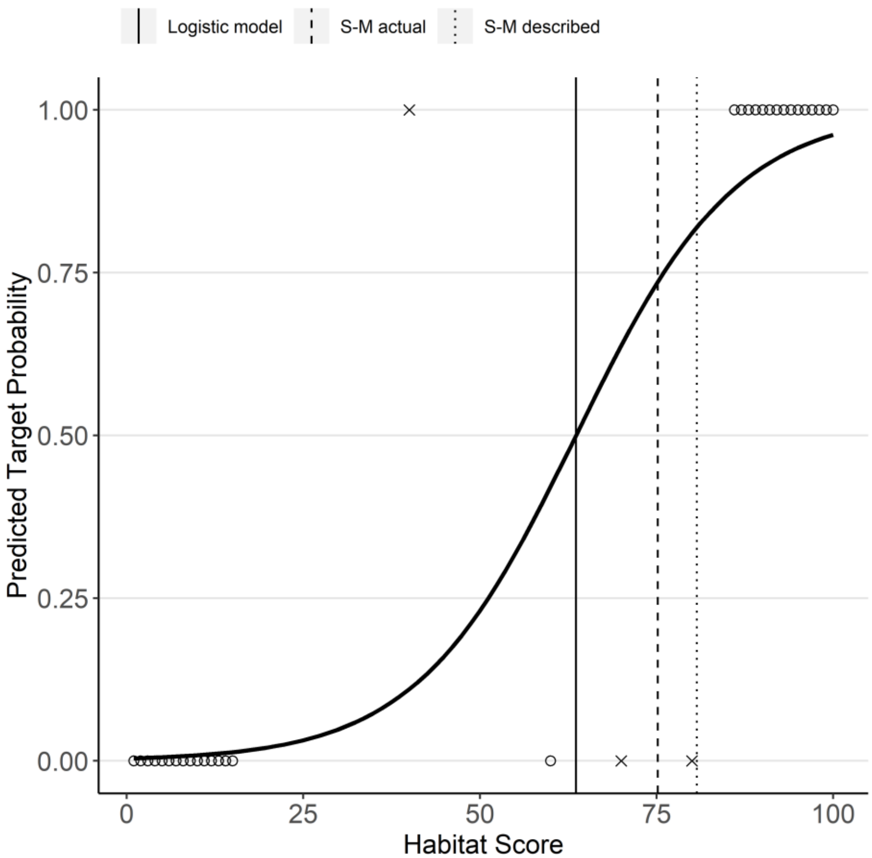

Fig. 1. Illustration of difference in classification thresholds applied to 34 artificial binary data points. Vertical lines from left to right indicate: 1. Maximum likelihood logistic model classification threshold p_{crit} = 0.5, i.e., the point at which the model estimates a trip is more likely targeted than non-targeted (solid line); 2. Threshold that minimizes difference between observed positive trips and predicted positive trips (Eq. 3) (dashed line); 3. Threshold that minimizes total misclassification rate (Eq. 2) (dotted line). Circles denote correct classification of observed values and x's denote incorrect classifications under p_{crit} = 0.5 assuming observed values are the truth. Notice the dashed and dotted line thresholds no longer rely on the logistic model and can be derived independently based on the observed values. The threshold as defined by Eq. (3) shifts the Eq. (2) threshold left to capture one additional negative trip to balance the positive trip outlier in upper left.

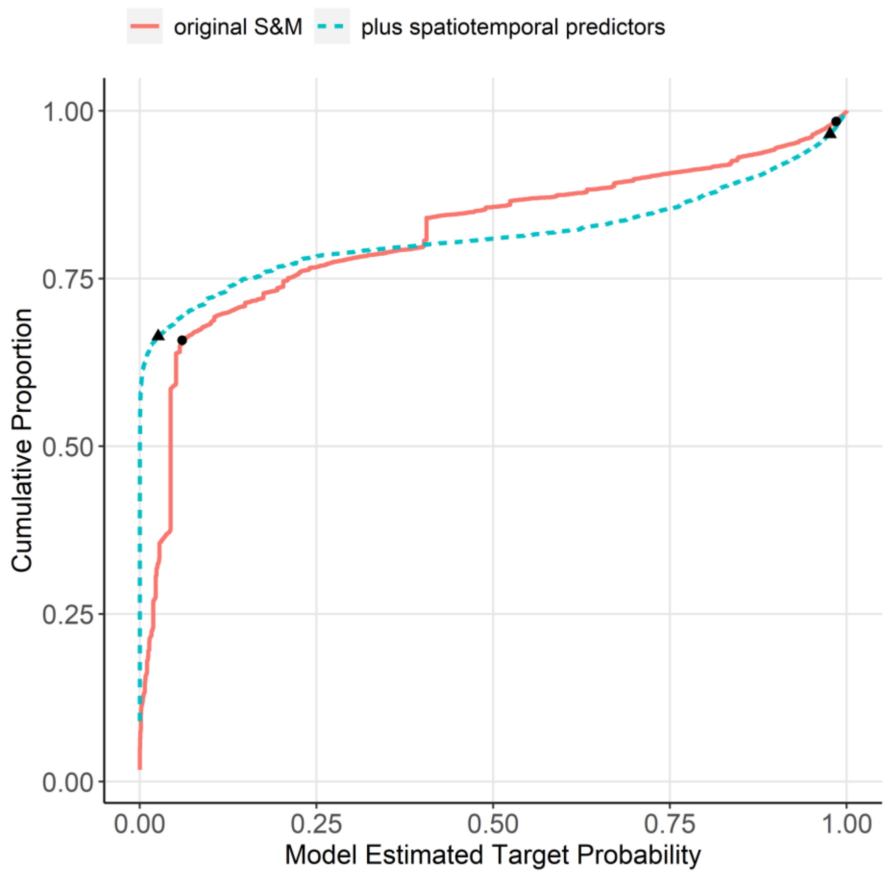

Fig. 2. Empirical cumulative distribution functions (eCDFs) of logistic model estimated target probabilities under original Stephens-MacCall model with only binary species predictors (solid red) and under revised model (last row Tbl. 1) with additional spatiotemporal predictors (dashed blue). Greater separation (vertical distance) along the edges with a flatter horizontal section is reflective of a more informative model. Markers represent points of diminishing returns to define trips that were very likely not targeted (at or below left marker, Eq. 4) or very likely targeted (at or above right marker, Eq. 5).

Model	***p***	**AIC**	**Δ AIC**
Intercept only	1	32,245	22,920
Original	176	16,369	7,044
Original + Depth	201	15,094	5,769
Original + Region	177	15,036	5,711
Original + Month	187	14,342	5,017
Original + Depth + Region	202	13,842	4,517
Original + Depth + Month	212	13,204	3,879
Original + Region + Month	188	11,947	2,622
Original + Depth + Region + Month	213	11,059	1,734
Original + Month + Depth * Region	238	10,733	1,408
Original + Depth + Region * Month	224	9,486	161
Original + Depth * Region + Region * Month	249	9,325	0

Tbl 1. Model selection table for logistic regressions fit to 2019 red snapper commercial logbook data to estimate predicted probabilities. "Original" denotes Stephens-MacCall model with only co-occurring species as predictors. Note AIC reductions in all cases with added spatiotemporal covariates. Interaction terms include all corresponding main effects.

	Original		**Revised**	
Non-targeted	22,159	1.9%	22,354	0.2%
Possibly targeted	10,974	48.2%	10,151	49.8%
Very likely targeted	538	95.9%	1,166	97.5%

Tbl. 2. Total 2019 commercial trips classified by likelihood of targeting red snapper according to empirical CDF thresholds of predicted probabilities from the original Stephens-MacCall model and final revised model with additional covariates. Percentages represent proportion observed trips with red snapper catch in each category. Note the relatively lower proportion of observed catch in trips classified as non-targeted and relatively higher proportion of observed catch in trips classified as very likely targeted under the revised model compared to the original model, indicating an improvement in classification performance.

	Original	**Revised**
Raw dataset	0.185	
"Non-targeted" trips removed	0.504	0.547
Weighted estimate (all trips)	0.622	***0.787***
Weighted estimate ("non-targeted" trips removed)	0.686	***0.791***
Trips with p_{pred} > 0.5	0.775	0.857

Tbl. 3. Comparison of CPUE according to various threshold criteria and weighting alternatives, with predicted probabilities estimated from original Stephens-MacCall model and revised model with additional covariates. Columns compare different threshold/weighting approaches under the same model and rows compare the same threshold/weighting approach under different models. Values range between 0 and 1 representing 0% successful trips and 100% successful trips, respectively. For example, the revised weighted estimate of 0.787 implies that a total of 7,912 trips potentially targeting red snapper occurred based on the 6,227 trips with observed red snapper catch. Note the sensitivity of CPUE to the different binary classification thresholds, lending support to trip weighting alternatives. Removing "non-targeted" trips had minimal impact on the revised model weighted estimate.

	Original	**Revised**
Common	4,200 (3,698)	
Unique	615 (35)	2,206 (1,791)
Total	4,815 (3,733)	6,406 (5,489)
$\boldsymbol{R^2_{McFadden}}$	0.503	0.726
Somers' *D*	0.863	0.964

Tbl. 4. Comparison of selected records between the original Stephens-MacCall model and revised model with additional covariates as applied to 2019 red snapper commercial logbook data (p_{crit} = 0.5 used for both). Values in parentheses denote the number of records with observed red snapper catch. Note superior performance of the revised model according to McFadden's pseudo R^2 and Somers' D fit metrics.