Using model-based inference to evaluate global fisheries status from landings, location, and life history data

James T. Thorson, Trevor A. Branch, and Olaf P. Jensen

Abstract: Assessing fishery collapses worldwide is hindered by the lack of biomass data for most stocks, leading to the use of landings-based proxies or the assumption that existing stock assessments are globally representative. We argue that the use of sparse assessments to evaluate fishery status requires model-based inference because assessment availability varies spatially and temporally, and we derive a model that extrapolates from assessment results to available landings, life history, and location data. This model uses logistic regression to classify stocks into different prediction bins and estimates the probability of collapse in each using cross-validation. Results show that landings, life history, and location are informative to discriminate among different probabilities of collapse. We find little evidence that regions with fewer assessments have a greater proportion of collapsed stocks, while acknowledging weak inferential support regarding regions with one or fewer assessments. Our extrapolation suggests that 4.5%–6.5% of stocks defined by landings data are collapsed, but that this proportion is increasing. Finally, we propose a research agenda that combines stock assessment and landings databases while overcoming limitations in each.

Résumé : Le manque de données sur la biomasse pour la plupart des stocks limite l'évaluation de l'effondrement des pêches à l'échelle mondiale, de sorte qu'il est nécessaire d'utiliser des données de substitution reposant sur les débarquements ou encore de présumer que les évaluations des stocks existantes sont représentatives de la situation à l'échelle planétaire. Dans le présent article, il est postulé que l'utilisation d'un nombre limité d'évaluations pour évaluer l'état des pêches nécessite le recours à l'inférence basée sur un modèle puisque la disponibilité des évaluations varie dans l'espace et dans le temps. Un modèle est en outre établi qui extrapole des résultats d'évaluation aux données disponibles sur les débarquements, le cycle de vie et l'emplacement. Ce modèle fait appel à la régression logistique pour classer les stocks selon différents compartiments de prédiction et estime la probabilité de l'effondrement de chaque stock à l'aide de la validation croisée. Les résultats montrent que les débarquements, le cycle de vie et l'emplacement constituent des renseignements utiles pour discriminer entre différentes probabilités d'effondrement. Peu de données probantes appuient la thèse voulant que les régions pour lesquelles moins d'évaluations sont disponibles aient une plus grande proportion de stocks effondrés, bien qu'il convienne de souligner la faiblesse du support inférentiel pour les régions pour lesquelles une seule évaluation ou moins est disponible. L'extrapolation suggère que de 4,5 % à 6,5 % des stocks définis par les données sur les débarquements sont effondrés et que cette proportion est en hausse. Enfin, un programme de recherche est proposé qui combine l'utilisation de bases de données sur l'évaluation des stocks et sur les débarquements afin de contourner les problèmes inhérents à chacun de ces outils.

[Traduit par la Rédaction]

Introduction

There is substantial debate over whether fisheries management has failed, both in scientific journals and popular media (Pauly 2009; Greenberg 2010; Hilborn 2011). This debate has practical importance, because scientific and public opinion shapes ongoing changes in fisheries management, legislation, and policy. However, insufficient data exist to estimate stock status, exploitation rates, or biomass trends for the majority of fish stocks worldwide.

As a result, two general strategies have been used to assess fishery status (and hence the quality of fisheries management) globally: global proxies and stock assessment data. Global proxies for fishery status have been previously identified, such as landings data, trophic level, or reconstructions of ecosystem productivity (Pauly 2010). These proxies provide valuable information regarding stocks and regions for which direct information about stock status is lacking, although the interpretation of such proxies may also be dis-

Received 22 June 2011. Accepted 31 January 2012. Published at www.nrcresearchpress.com/cjfas on 20 March 2012.
J2011-0273

Paper handled by Associate Editor Terrance Quinn.

J.T. Thorson and T.A. Branch. School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, WA 98195-5020, USA.
O.P. Jensen. Institute of Marine and Coastal Sciences, Rutgers University, 71 Dudley Road, NB, NJ 08901-8525, USA.

Corresponding author: James T. Thorson (e-mail: JimThor@uw.edu).

puted or equivocal. The use of landings data as a proxy for stock biomass famously led to the prediction that current trends would result in the collapse of all fish stocks by 2048 (Worm et al. 2006).

By contrast, other research has sought to assess global status by compiling stock assessment estimates of biomass, exploitation rates, and other detailed data (Worm et al. 2009; Branch et al. 2011; Ricard et al., in press). This research supports the conclusion that fishing mortality rates for the majority of stocks in the developed world are at or below the rate that should result in maximum sustainable yield (Worm et al. 2009), and on average, biomass of exploited stocks has stabilized in the last two decades (Hutchings et al. 2010). However, less is known about the stock status of fishes in developing regions, and there are reasons to be concerned about the status of these stocks (e.g., the global increase and redistribution of fishing effort towards developing regions; Worm et al. 2009; Anticamara et al. 2011).

Proxies are subject to potentially large errors because they may not be consistently indicative of fishery health; for example, decreases in landings may not indicate stock collapse (Mullon et al. 2005; Carruthers et al. 2012). In particular, landings-based proxies may systematically underestimate stock collapse during early stages of fishery development, while overestimating collapse for regions that are subject to catch controls (Branch et al. 2011; Daan et al. 2011). Conversely, relying solely on stock assessment results to support inference to global fishery status can lead to errors because of the implicit assumption that assessed stocks are representative of all stocks worldwide. Specifically, assuming that assessments are regionally or globally representative fails to account for different sampling intensities between developed and developing regions, commercial and noncommercial species, and other systematic changes in stock assessment availability across space and time. Although previous studies have compared stock assessment- and landings-based proxies for stock collapse (de Mutsert et al. 2008; Branch et al. 2011; Pinsky et al. 2011), no previous study has presented results that simultaneously incorporate both types of information to estimate global fishery status.

In this study, we argue that the use of opportunistically collected stock assessment data to analyze global fisheries status requires model-based inference to account for data availability that varies within and between regions and over time. We assess whether globally available landings, life history, and location information are sufficiently informative to allow model-based predictions regarding stocks for which single-species assessments are not available. We then present an extrapolation model that uses stock assessment results to estimate parameters for a model that predicts the probability of collapse for those stocks in the Food and Agriculture Organization of the United Nations (FAO) global landings database. We use this model to estimate the proportion of stocks that are collapsed globally, as well as between "data-rich" and "data-poor" regions, and compare results with the commonly used landings-based proxy for stock collapse to demonstrate the impact of using stock assessment, life history, and location information in addition to FAO landings data. We believe this model represents an initial attempt at a potentially broad class of models that accounts for species with sparse fishery or biological data while addressing global-scale issues such as global stock status and fishery management success.

Materials and methods

We developed a modified logistic regression model (which we term an "assessment-calibrated extrapolation model" or ACEM) that treats assessment estimates of stock collapse as a response variable, which is predicted by statistics extracted from landings data, life history, and location information. The model is cross-validated through comparison with stock assessment data. While we acknowledge that stock assessments will not always accurately identify instances of collapse (see Carruthers et al. 2012 for an example using surplus production models), we believe that these stock assessments represent the most reliable source of information regarding single-species stock status and instances of stock collapse.

The ACEM involved four steps: (1) estimate parameters for a logistic regression model with mixed effects that predicts stock assessment collapse using landings statistics, life history data, and location to classify stocks into five prediction bins, ranging from least to most likely to be collapsed; (2) use cross-validation (as detailed below) within the stock assessment database to estimate the conditional probability of "true" collapse in each of these five bins; (3) extrapolate the logistic regression to classify stocks in the FAO landings database into the same five bins; and (4) apply the cross-validation results (step 2) to estimate the global proportion of collapsed stocks.

Available data

We extracted assessment estimates for 232 stocks (including finfish and invertebrates, with average duration of 35 years) from the RAM Legacy Stock Assessment Database (Ricard et al., in press). These data include landings and either annual estimates of spawning biomass (SSB) and spawning biomass at maximum sustainable yield (SSB_{msy}) or total biomass (TB) and total biomass at maximum sustainable yield (TB_{msy}). We coded all occurrences when SSB or TB dropped below 20% of their maximum sustainable yield (MSY) levels as collapsed, as in Worm et al. (2009) and Branch et al. (2011), and treated this as a response variable that is explained by landings, location, and life history data. Landings data were extracted from the FAO FishStat database (Food and Agriculture Organization of the United Nations 2010*a*), and landings data for each reported taxon (including invertebrates) in a FAO major fishing area were treated as a "stock". Analyses of biomass and landings time series were restricted to stocks with a total catch summed across 1950–2008 greater than 10 000 t to match the methods used in Worm et al. (2006); this restriction resulted in 1886 stocks with FAO landings time-series data. Landings data were processed with a 3-year moving average filter to reduce interannual variability that is not likely to be informative about fishery collapse (following Pinsky et al. 2011), and smoothed landing data were used to calculate relative catch statistics (i.e., catch expressed as the percentage of previous maximum catch). Effects of previous catch levels were incorporated by using lagged relative catch for 1, 2, and 3 years prior to a given year as a predictor. Including lagged effects precluded predictions of stock collapse at the beginning of

Table 1. Variables used in cross-validation analysis and the assessment-calibrated extrapolation model (ACEM), classified by data type (biomass, descriptive, landings, life history, or location) and variable type (response, predictor), and indices used in the analysis.

Name	Symbol	Data type	Variable type
Variables			
Biomass-based collapse for stock i in year y	$\mathrm{BC}_{i,y}$	Biomass	Response
Intercept	—	Descriptive	Predictor
Calendar year for stock i in year y	$T_{i,y}$	Descriptive	Predictor
Calendar year squared for stock i in year y	$T_{i,y}^2$	Descriptive	Predictor
Relative catch (smoothed) for stock i in year y	$\mathrm{RC}_{i,y}$	Landings	Predictor
Relative catch (smoothed + lag 1) for stock i in year y	$\mathrm{RC}(1)_{i,y}$	Landings	Predictor
Relative catch (smoothed + lag 2) for stock i in year y	$\mathrm{RC}(2)_{i,y}$	Landings	Predictor
Relative catch (smoothed + lag 3) for stock i in year y	$\mathrm{RC}(3)_{i,y}$	Landings	Predictor
Catch variability for stock i in year y	$\mathrm{SD}_{i,y}$	Landings	Predictor
Maximum recorded length for stock i in year y	$L_{i,y}$	Life history	Predictor
Trophic level for stock i in year y	$\mathrm{TL}_{i,y}$	Life history	Predictor
FAO major fishing area for stock i in year y	$F_{i,y}$	Location	Predictor
Number of unique assessments included in the RAM Legacy Stock Assessment Database in FAO region f	N_f	Location	Predictor
Indices			
Stock ID	i		
Year	y		
FAO major fishing area	f		
Cross-validation replicate	j		
Cross-validation bins	b		

landings history (i.e., for years 1950–1952). Variability in catch was also used as a predictor of stock collapse and was calculated as the standard deviation of relative catch for all years prior to a given year (Table 1). Relative catch was set to one and variability was set to zero in the FAO landings database prior to exploitation.

Trophic level and maximum recorded length were obtained from FishBase (Froese and Pauly 2010). They were used as life history information for each stock because preliminary analysis indicated that stocks in the RAM database have on average higher trophic level or greater length than those in the FAO database, implying that an extrapolation model should explicitly account for differences in average life history trait between databases. For taxa not recorded to species level, or where species-specific trophic-level estimates were not available, the mean trophic level of all species in the most closely related taxonomic grouping was used (within higher-level taxa like genus, family, order, class, or phylum). FishBase trophic level values for each species were selected primarily based on published studies reporting the percentage in mass or volume of different food items in the diet and secondarily on reports of individual food items if necessary (Branch et al. 2010).

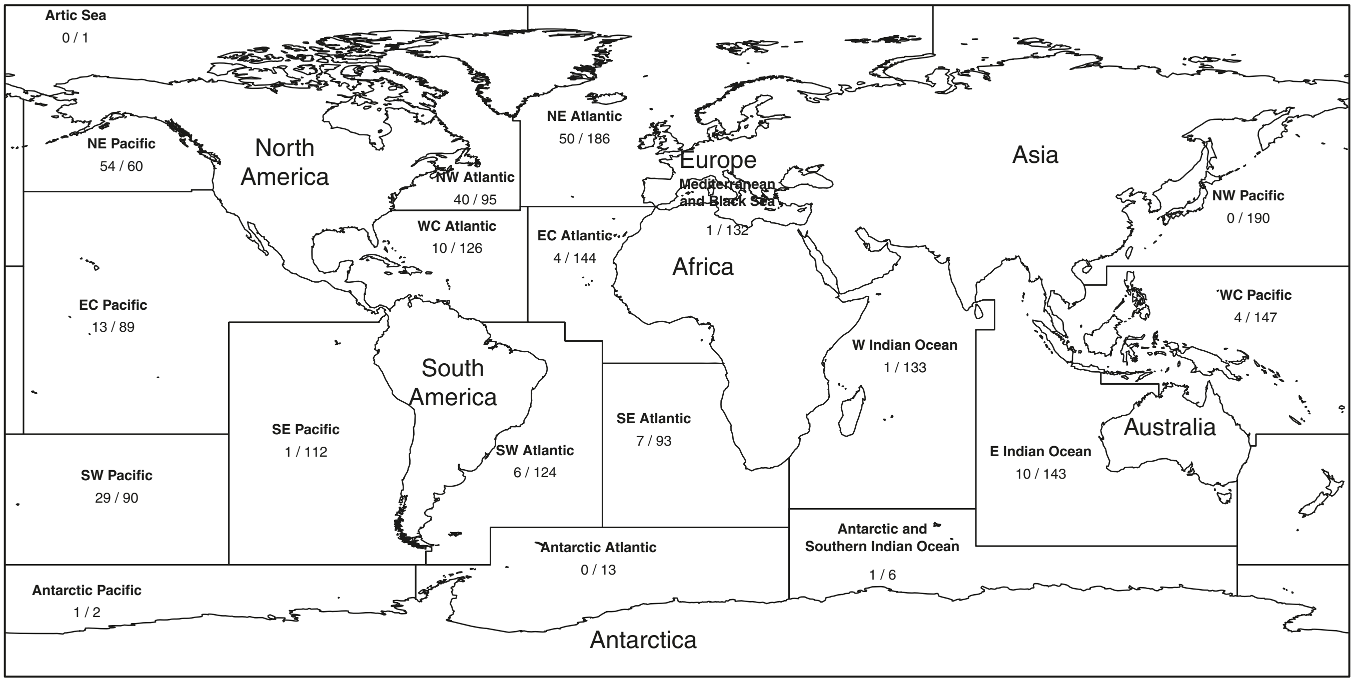
Geographic region was recorded as the FAO major fishing area for each stock, for example the northwest Atlantic. Additionally, the number of unique stock assessments in the RAM Legacy Stock Assessment Database (i.e., after excluding update assessments for the same stock) was selected as a predictor variable representing the degree to which a region has formal single-species fisheries management in place. This variable was used instead of the number of assessments in each year to avoid confounding with calendar year, which was also included as a predictor variable. Inspection of data availability by region for FAO landings and RAM Legacy Stock Assessment databases (Fig. 1) shows that stock assessment data are available primarily for North America (eastern central Pacific, northeast Pacific, and northwest Atlantic) and Europe (northeast Atlantic), although the major fishing area including New Zealand and southeast Australia also had a large proportion of assessed stocks. We will refer to the northeast (NE) Atlantic, northwest (NW) Atlantic, NE Pacific, and southwest (SW) Pacific as "data-rich" regions and other regions as relatively "data-poor." We acknowledge that ongoing expansion of the RAM Legacy Stock Assessment Database will change the relative proportion of available stock assessments in each region, but do not believe it will qualitatively change which regions have relatively few or many stock assessments. Sixteen of the 19 FAO major fishing areas had at least one stock assessment, while nine had greater than five assessments.

Fitting the ACEM to stock assessment data

Logistic regression stage

The first stage of the ACEM model-fitting process involves a logistic regression model with mixed effects, where stock collapse (identified using biomass data) is treated as a response variable and landings, life history, and location variables are used as predictors (Table 1). In this stage, landings data are extracted from the RAM Legacy Stock Assessment Database to ensure that the spatial scale of landings time-series data corresponds to the stock definition used in the estimated biomass time series. In this logistic regression model, the probability that a given stock i is collapsed in year y is

Fig. 1. Map of Food and Agriculture Organization of the United Nations (FAO) major fishing areas, with FAO name, followed by the number of RAM Legacy stock assessments and the number of species with FAO landings data in each region.



estimated using a mixed-effects logistic regression model (eqs. 1–4):

$$(1)\qquad p_{t,y} = \Big(1+\exp\Big\{-\Big[\widehat{\beta}_I + \widehat{\beta}_T T_{i,y} + \widehat{\beta}_{T^2} T^2_{i,y} + \widehat{\beta}_{\mathrm{RC}}\mathrm{RC}_{i,y} + \widehat{\beta}_{\mathrm{RC(1)}}\mathrm{RC}(1)_{i,y} + \widehat{\beta}_{\mathrm{RC(2)}}\mathrm{RC}(2)_{i,y} + \widehat{\beta}_{\mathrm{RC(3)}}\mathrm{RC}(3)_{i,y} + \widehat{\beta}_{\mathrm{SD}}\mathrm{SD}_{i,y} + \widehat{\beta}_L L_i + \widehat{\beta}_{\mathrm{TL}}\mathrm{TL}_i + \sum\nolimits_f \widehat{\beta}_f I(F_i=f) + \sum\nolimits_f \widehat{\beta}_{\mathrm{RC},f} I(F_i=f)\mathrm{RC}_{i,y}\Big]\Big\}\Big)^{-1}$$

$$(2)\qquad \overline{\beta}_f = \sum\nolimits_f \widehat{\beta}_N I(F_i=f)N_f$$

$$(3)\qquad \overline{\beta}_{\mathrm{RC},f} = \sum\nolimits_f \widehat{\beta}_{\mathrm{RC},N} I(F_i=f)N_f$$

$$(4)\qquad L = \int\limits_{\widehat{\beta}_f}\int\limits_{\widehat{\beta}_{\mathrm{RC},f}} \left\{ \frac{1}{\sqrt{2\pi\widehat{\sigma}^2_{\beta_f}}}\exp\left[\frac{-(\widehat{\beta}_f-\overline{\beta}_f)^2}{2\widehat{\sigma}^2_{\beta_f}}\right]\frac{1}{\sqrt{2\pi\widehat{\sigma}^2_{\beta_{\mathrm{RC},f}}}} \times \exp\left[\frac{-(\widehat{\beta}_{\mathrm{RC},f}-\overline{\beta}_{\mathrm{RC},f})^2}{2\widehat{\sigma}^2_{\beta_{\mathrm{RC},f}}}\right]\prod_{i=1}^{n_{\mathrm{stocks}}}\prod_{y=1}^{n_{\mathrm{years}}} p_{i,y}^{\mathrm{BC}_{i,y}} \times (1-p_{i,y})^{1-\mathrm{BC}_{i,y}} \right\} \mathrm{d}\widehat{\beta}_f\, \mathrm{d}\widehat{\beta}_{\mathrm{RC},f}$$

where $p_{t,y}$ is the estimated probability of collapse for stock i in year y; variables are defined in Table 1; estimated coefficients $\widehat{\beta}$ are defined with subscripts that correspond to these parameters and all estimated parameters are denoted with a hat symbol; $I(F_i = f)$ is an indicator variable that equals 1 if stock i is in FAO region f and 0 otherwise; $\widehat{\sigma}^2_{\beta_f}$ and $\widehat{\sigma}^2_{\beta_{\mathrm{RC},f}}$ are the variance for $\widehat{\beta}_f$ and $\widehat{\beta}_{\mathrm{RC},f}$ random effects, which have expected values $\overline{\beta}_f$ and $\overline{\beta}_{\mathrm{RC},f}$, respectively; $\mathrm{BC}_{i,y}$ are instances of collapse identified from biomass data, coded as 1 for stocks where $\mathrm{TB}_{i,y} < 0.2\mathrm{TB}_{\mathrm{MSY}}$ or $\mathrm{SSB}_{i,y} < 0.2\mathrm{SSB}_{\mathrm{MSY}}$ and 0 otherwise; and eq. 4 is the marginal likelihood that is maximized, which includes integrals across all random effects.

In the ACEM, the FAO major fishing area for each stock was used as a grouping variable for two random effect coefficients per region: the intercept (β_f; i.e., collapse probabilities in the absence of other effects; eq. 2) and the slope for relative catch ($\beta_{\mathrm{RC},f}$; i.e., the impact of changes in relative landings on the probability of collapse; eq. 3). The intercept was treated as a random effect (eq. 2) because this proportion is likely to vary among regions because of unmeasured management actions that affect the probability of collapse for all stocks in a region. The slope for relative catch was treated as a random effect (eq. 3) because the management measures used in some FAO regions are likely to cause decreases in relative catch to be less or more informative about stock collapse than in other regions. Using random effects also allows for extrapolation to the 3 of 19 FAO major fishing areas for which we do not have data to train the extrapolation model (i.e., with no observations in the RAM Legacy Stock Assessment Database).

Additionally, the number of assessed stocks (N_f) and its interaction with relative catch ($\mathrm{RC}_{i,y} \times N_f$) are included as variables that model the expected value for regional random effects (e.g., $E[\widehat{\beta}_f] = \sum_f \widehat{\beta}_N I(F_i=f)N_f$ and $E[\widehat{\beta}_{\mathrm{RC},f}] = \sum_f \widehat{\beta}_{\mathrm{RC},N} I(F_i=f)N_f$ in eqs. 2–3). These variables and associated hyperparameters $\widehat{\beta}_N$ and $\widehat{\beta}_{\mathrm{RC},N}$ allow the expected value for regional effects to diverge from the worldwide value even for regions with no data to inform their

Table 2. Model coefficients are classified by effect type (fixed or random) and selection type (either selected using the stepwise model selection using Akaike's information criterion (AIC) or included a priori).

	Term in eq. 1	Effect type	Selection type
Descriptive			
Intercept	$\widehat{\beta}_I$	Fixed	A priori
Calendar year	$\widehat{\beta}_T$	Fixed	AIC
Calendar year squared	$\widehat{\beta}_{T^2}$	Fixed	AIC
Landings			
Relative catch (smoothed)	$\widehat{\beta}_{\mathrm{RC}}$	Fixed	A priori
Relative catch (smoothed + lag 1)	$\widehat{\beta}_{\mathrm{RC(1)}}$	Fixed	AIC
Relative catch (smoothed + lag 2)	$\widehat{\beta}_{\mathrm{RC(2)}}$	Fixed	AIC
Relative catch (smoothed + lag 3)	$\widehat{\beta}_{\mathrm{RC(3)}}$	Fixed	AIC
Catch variability	$\widehat{\beta}_{\mathrm{SD}}$	Fixed	AIC
Life history			
Maximum recorded length	$\widehat{\beta}_L$	Fixed	A priori
Trophic level	$\widehat{\beta}_{\mathrm{TL}}$	Fixed	A priori
Location			
Region-specific deviation from global intercept	$\widehat{\beta}_f$	Random	A priori
Region-specific deviation from global relative catch	$\widehat{\beta}_{\mathrm{RC},f}$	Random	A priori
Effect of number of assessed stocks in an FAO region (in 2008)	$\widehat{\beta}_N$	Fixed	AIC
Effect of number of assessed stocks in an FAO region (in 2008) on the relative catch	$\widehat{\beta}_{\mathrm{RC},N}$	Fixed	AIC

value. Instead, the expected value for regional effects varies as a linear function of the number of assessments in each region (i.e., $\widehat{\beta}_f$ and $\widehat{\beta}_{\mathrm{RC},f}$ will be shrunk to $\sum_f \widehat{\beta}_N I(F_i = f)N_f$ and $\sum_f \widehat{\beta}_{\mathrm{RC},N} I(F_i = f)N_f$, respectively), allowing model-based inference regarding the effect of Intercept and $\mathrm{RC}_{i,y}$ variables for regions with no available assessments. For example, the Intercept random effect for a region without assessments reverts to an expected value characteristic of such regions, as estimated from the relationship between N_f and Intercept for regions with assessments.

Some coefficients are included a priori in this model, while others are selected using stepwise model selection and the Akaike information criterion (AIC, Akaike 1974). Specifically, we include an intercept, the relative catch (without lag), all life history variables, and all random effects a priori, but select among the remaining variables, including statistics extracted from landings data and the number of assessments per region using stepwise AIC model selection (Table 2). These selected coefficients and estimated parameter values will be used later for extrapolation to the FAO landings database. Fixed effects are selected using stepwise model building and AIC after including random effects. Parameter values for this logistic regression model with mixed effects are estimated by fitting to stock assessment data from the RAM Legacy Stock Assessment Database (Ricard et al., in press) using the "lme4" package (Bates and Maechler 2009) in the R statistical computing environment (R Development Core Team 2009). Stepwise selection is conducted using maximum marginal likelihood, while the selected model is then re-run using restricted maximum likelihood as recommended by Zuur et al. (2009).

Cross-validation stage

The second stage of the ACEM model-fitting process aims to account for model misspecification (i.e., overdispersion, nonlinear effects, and the influence of missing variables). It does this by estimating the conditional probability of stock collapse given predictions from the logistic regression model using cross-validation within the set of stocks with data in the RAM Legacy Stock Assessment Database. In each cross-validation replicate, 25 randomly selected stocks were set aside as a "testing set", and all data from these stocks were excluded (although the number of assessments per FAO region N_f was unchanged); the remaining data were a "training set." The logistic regression model is fitted to the training set, using stepwise AIC model selection to select among possible coefficients. Estimated parameter values for these selected coefficients are then used to predict stock collapse for a randomly chosen year for each of the 25 stocks in the testing set. Predicted probability of collapse ($p_{t,y}$ for a randomly chosen year y for stock i in the testing set, ranging continuously from zero to one) and true stock status (biomass collapse, $\mathrm{BC}_{i,y}$) for these years of the testing set are recorded, and the process is repeated 4000 times (yielding 100 000 predictions). These ACEM predictions of collapse probability (ranging from 0 to 1) were then binned into five prediction "bins" (0–0.2, 0.2–0.4, 0.4–0.6, 0.6–0.8, 0.8–1), representing different estimated probabilities of stock collapse. The true stock status for each binned prediction was used to calculate the proportion of stocks in each predictive bin that was actually collapsed. The proportion of stocks assigned to each prediction bin that are actually collapsed is treated as a nonparametric estimate of the conditional probability of stock collapse given a logistic regression prediction ($\widehat{p}_b$). This process was done because model misspecification and missing variables can cause the true probability of stock collapse to diverge from the parametric form assumed by conventional logistic regression (similar to the effect seen in Carruthers et al. 2012).

Cross-validation was also used to calculate the area under the received operator characteristics curve (AUC), which is a threshold-independent metric of predictive performance for models such as the ACEM that predict a binary outcome. Specifically, the estimated probability of collapse (p_j) and true status (BC_j) for each cross-validation observation j was compiled. Next, the sensitivity

$$\left[\frac{\sum_{j=1}^{n_{\text{crossval}}} I(p_j \geq p_{\text{threshold}})\mathrm{BC}_j}{\sum_{j=1}^{n_{\text{crossval}}} \mathrm{BC}_j}\right]$$

and specificity

$$\left[\frac{\sum_{j=1}^{n_{\text{crossval}}} I(p_j \leq p_{\text{threshold}})(1-\mathrm{BC}_j)}{\sum_{j=1}^{n_{\text{crossval}}} (1-\mathrm{BC}_j)}\right]$$

were calculated for 101 possible thresholds $p_{\text{threshold}}$ ranging from 0 to 1, and the AUC composed of linear interpolation between the sensitivity and specificity for each possible threshold value was calculated (Hosmer and Lemeshow 2000). This AUC value could range from 0.5 (i.e., no predictive power) to 1 (i.e., perfect predictive power).

Extrapolating to global status

After fitting the ACEM to stock assessment data, we used estimated logistic regression parameters and cross-validation estimates to extrapolate to the global FAO landings database. In this extrapolation stage, data from the FAO landings database is used to classify all stocks in the database in each year into one of the five prediction bins, and the number n_b of stocks in each bin b is recorded. The proportion collapsed in the FAO landings database was estimated by the ACEM as the sum across bins of expected number of collapses in each bin, divided by the total number of stocks. Conditional probability of collapse $\widehat{p}_b$ for each bin b was estimated from cross-validation using only RAM Legacy Stock Assessment data (i.e., Table 3), while the bin assignment for each stock in the FAO landings database in a given year was accomplished using the ACEM. Each bin was treated as a binomial distribution with an estimated probability of collapse $\widehat{p}_b$ and the number of FAO stocks assigned to that bin n_b (eq. 5).

$$(5)\qquad E[\text{Prop. collapsed}] = E\left[\frac{1}{\sum_{b\in(\text{bins})} n_b} \times \sum_{b\in(\text{bins})} \text{binomial}(\widehat{p}_b, n_b)\right] = \frac{1}{\sum_{b\in(\text{bins})} n_b} \sum_{b\in(\text{bins})} \widehat{p}_b n_b$$

The variance for the proportion collapsed was then calculated using Monte Carlo simulation of independent binomial trials for each bin (and assuming a fixed $\widehat{p}_b$).

Table 3. Cross-validation results for the assessment-calibrated extrapolation model (ACEM).

		Cross-validation assignments				
Stock assessment status	No. of stock-years in RAM Legacy Database	Not collapsed ($0.0 \leq p_{i,y} < 0.2$)	Likely not collapsed ($0.2 \leq p_{i,y} < 0.4$)	Unknown ($0.4 \leq p_{i,y} < 0.6$)	Likely collapsed ($0.6 \leq p_{i,y} < 0.8$)	Collapsed ($0.8 \leq p_{i,y} \leq 1.0$)
Not collapsed	7 766	78 143	6 944	4 328	2 896	2 477
Collapsed	431	2 078	598	597	714	1 225
Derived parameters						
Proportion actually collapsed in RAM Legacy Database	—	0.026	0.079	0.121	0.198	0.331
Proportion of collapsed stocks in each bin	—	0.399	0.115	0.115	0.137	0.235

Note: The ACEM was used to categorize data for each year and stock from the RAM Legacy Stock Assessment Database into one of five categories representing different probability of stock collapse. The first two rows show the number of stock-years in the RAM Legacy Stock Assessment Database that are actually collapsed and not collapsed for each predictive bin, with the third row showing the proportions of stocks in each predictive bin that are actually collapsed. The final row shows the proportion of collapsed stocks in each predictive bin.

We compared the ACEM with another widely used method for analyzing global landings data, which classifies any stock as collapsed when catches have dropped below 10% of the maximum catch seen in any year (Worm et al. 2006; Costello et al. 2008; Pauly and Froese 2012). A 95% confidence interval for this method was calculated in each year by applying the nonparametric bootstrap using three steps: (*i*) sampling stocks with replacement repeatedly from the set of stocks that were available in any given year; (*ii*) replicating the analysis method for each bootstrap sample; (*iii*) aggregating all bootstrap estimates for a given year and calculating the 2.5% and 97.5% quantiles.

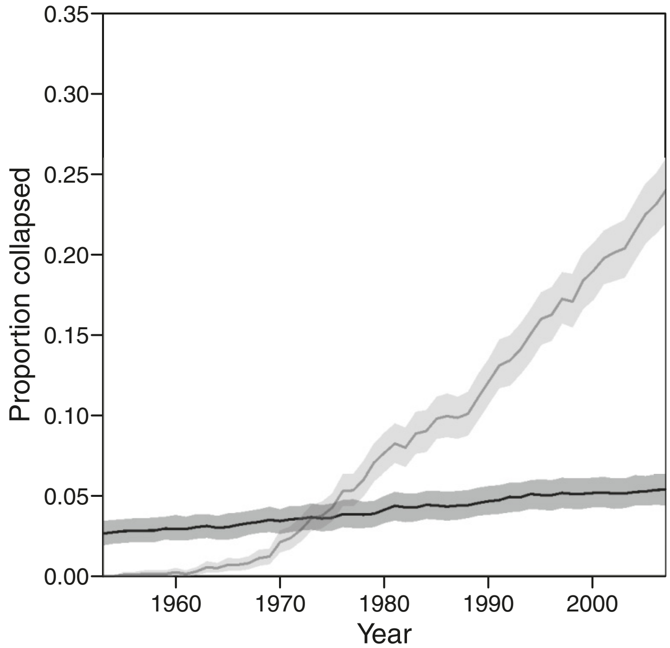
Results and discussion

Cross-validation results (Table 3) for the five prediction bins show that the ACEM classifies 40% of collapses in the RAM Legacy Stock Assessment Database into the "not collapsed" bin, but that the ACEM is correct 97% of the time when classifying a stock to this bin. By contrast, the ACEM is only correct for 20% and 33% of stocks that it classifies in the "probably collapsed" or "collapsed" bin, to which 14% and 24% of true collapses are classified, respectively. The frequency of true collapse increases monotonically from 3% to 33% as predicted status goes from "not collapsed" (Bin 1) to "collapsed" (Bin 5). The divergence between theoretical and cross-validation predictive accuracy for the higher bins (Bins 3–5) motivates the present two-stage extrapolation process, because the accuracy of each bin diverges from the form that is assumed by logistic regression. This divergence is likely caused by the "global support" of each observation implied by conventional logistic regression, where each observation has leverage regarding the estimated probability of collapse for all other observations (Hastie et al. 2009). The result of low predictive accuracy regarding collapsed stocks is not surprising, given that collapse in any given year is a rare event, and relevant economic, management, and biological information are missing. The large difference in true probability of collapse between bins (i.e., 3% in the "not collapsed" bin vs. 33% in the "collapsed" bin) implies that globally available proxies do provide substantial (though imperfect) discrimination between collapsed and not-collapsed stocks. Thus, cross-validation results imply that landings, life history, and location are informative about the prevalence of collapse despite this relatively low predictive accuracy for individual collapsed stocks.

The AUC was 0.75 for the ACEM predictions of stock collapse. This AUC level is characterized as acceptable by Hosmer and Lemeshow (2000), which implies that landings, life history, and location information is informative (if imperfect) to predict instances of stock collapse. This result forms the basis of subsequent efforts to use globally available information to extrapolate to the proportion of stocks that are collapsed worldwide.

When the ACEM was applied to all stocks globally, it estimated that the proportion of collapsed stocks worldwide has increased slowly but steadily since the 1950s, reaching a current rate of 5.4% (Fig. 2). By contrast, our replication of results using the landings-based method shows a current collapse rate of 24%. Although the increasing trend is consistent in both methods, the absolute level is much lower using the ACEM. The ACEM estimates a magnitude of collapse that is intermediate between the values of 14% estimated by Worm et al. (2009) or 7%–13% estimated by Branch et al. (2011) and the 4% of stocks classified as collapsed or recovering from collapse by the FAO from their most recent global assessment of stock status (Food and Agriculture Organization of the United Nations 2010*b*). The ACEM estimates an increasing trend in stock collapse, similar to the trend in the proportion of collapsed stocks in the RAM Legacy Stock Assessment Database (Worm et al. 2009). This estimated increasing trend conflicts with the stabilization of total biomass in the 1990s seen in the assessment database (Worm et al. 2009; Hutchings et al. 2010) and the reduction in collapsed stocks in the most recent FAO status report (Food and Agriculture Organization of the United Nations 2010*b*).

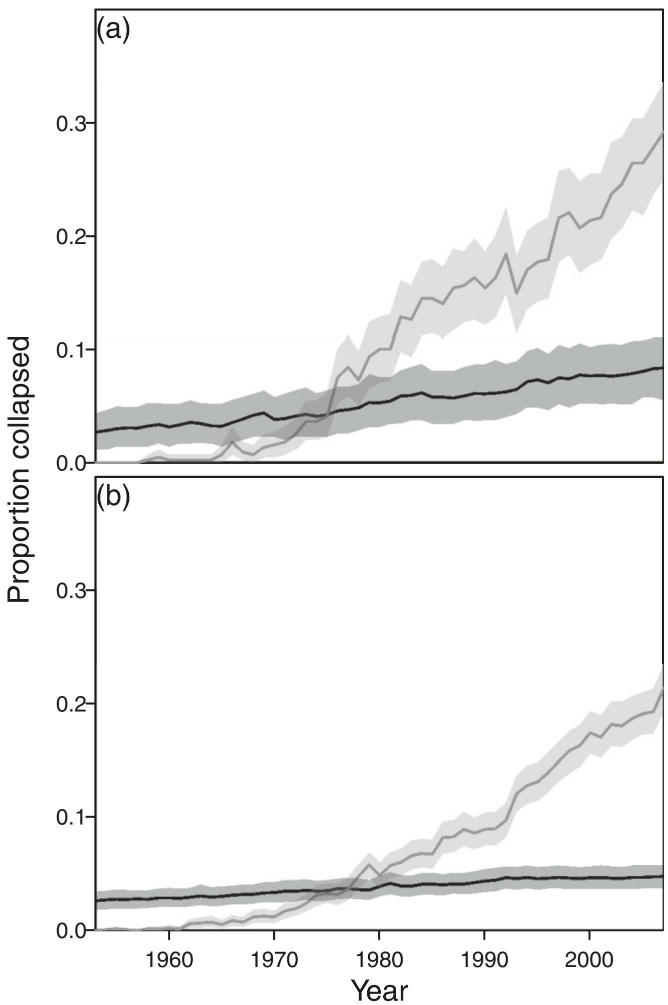
Fig. 2. Proportion of global stocks estimated to be collapsed using the ACEM (dark line) and from the landings-based method (light line); shading indicates 95% confidence intervals.



The ACEM also estimates a greater proportion of collapsed stocks in areas with a large number of available stock assessments (NE Pacific, NW Atlantic, NE Atlantic; 8.5% collapse) compared with all other FAO major fishing areas (4.8% collapse; Fig. 3). This result is consistent with the landings-based method, which also finds more collapses in data-rich regions (25%–33%) than in data-poor regions (19%–23%).

Inspection of estimated fixed effects (Table 4) is also informative regarding model performance as well as which parameters were associated with collapse. Relative catch in the current year has a large negative impact on the probability of collapse, implying that stocks where landings are greatly reduced have a much higher probability of being collapsed. Variability in catch is selected by AIC and has a positive sign, but a Wald test does not show statistical significance. Among life history traits, high trophic level species are less likely to be collapsed, and large-bodied fishes are slightly more likely to experience stock collapse; the latter estimate

Fig. 3. Proportion of stocks listed in the FAO catch database predicted to be collapsed in years 1953–2007 for data-rich regions (NE Pacific, NW Atlantic, NE Atlantic, SW Pacific) and data-poor regions (all others), for landings-based collapse (grey line and light grey interval) and the ACEM predictions (black line and dark grey interval), with confidence intervals from bootstrap and Monte Carlo simulation, respectively.



runs counter to previous analysis by Pinsky et al. (2011), perhaps because our multiple regression approach simultaneously estimates the effect of all variables in the model. Neither the number of assessed species in a region nor its interaction with relative catch was selected by AIC. The estimated random effects (Table 5) for the intercept are positive for the NW and NE Atlantic and negative for the NE Pacific (among other regions), implying that stocks in the NW and NE Atlantic have a greater probability of stock collapse than in the NE Pacific. Similarly, random effects for relative catch are negative for the NE Atlantic and positive for the NW Atlantic and NE Pacific, implying that decreases in landings are more informative about stock collapse in the NE Atlantic and less informative about stock collapse in the NW Atlantic and NE Pacific.

Research agenda for future work

For many researchers, the idea of "extrapolating" from assessment to global landings data may seem suspect. However, extrapolation is already implied by using assessment data to support statements about regional or global fishery status, which includes stocks for which data are not available. Such statements would be supported if assessment data were collected in a randomized design, but instead they are collected opportunistically (i.e., in a nonrandomized design) where sampling intensities differ spatially and over time. Problems arising from opportunistic and unbalanced data are already widely acknowledged in fisheries science when analyzing fishery catch and effort data (Walters 2003), leading to the use of "index standardization" models when estimating trends in stock abundance (Maunder and Punt 2004). These index standardization models are designed to account for unbalanced sampling designs in a manner that is analogous to the ACEM. However, the role of model-based inference when using opportunistic and unbalanced data has not been previously acknowledged in studies that assess global fishery status using stock assessment results.

For other researchers, the use of stock assessment results at all will be suspect, given that these data can potentially represent a biased sample from the most managed regions of the world. To account for this, we include region as a grouping variable for random effects (for which we have at least one assessment for 16 of the 19 FAO major fishing areas), model the expected value of random effects as a function of the number of existing stock assessments as a proxy for single-species management effort, and also account for systematic differences in landings history and life history by including these variables as covariates. Covariates are notably lacking for economic or multispecies management effort, and the analysis would be improved by additional covariates at the stock or region level that account for these differences among stocks and regions.

Despite our best efforts in this study, we believe that this work is still a preliminary step in the development of global extrapolation models, which use the opportunistic data that are available on a global scale for evaluating fishery questions. Future research could evaluate many possible directions and concerns as listed below.

Different stock definitions

In this study, we used management definitions of stocks implied by single-species stock assessments for model-fitting, and then extrapolated to stocks defined on a FAO major fishing area spatial scale. This has implications for the different definitions of landings data used in the RAM Legacy Stock Assessment (i.e., model-fitting and cross-validation) and FAO (i.e., extrapolation) model components, although we do not think that spatial-scale differences will cause systematic differences in the interpretation of relative catch time-series data as used in this model. This difference in spatial scale will also affect the definition of "stock collapse" between RAM and FAO model components, and the effect of this is difficult to evaluate without public access to landings data defined at a fine spatial scale.

Structured cross-validation

Within our model-fitting, the second stage was a cross-validation step that was intended to account for model misspecification. Future efforts could improve this cross-

Table 4. Fixed effects (either included a priori or selected using stepwise AIC model building) estimated using restricted maximum likelihood when fitting the logistic regression (eq. 1) to all RAM Legacy Stock Assessment data, along with the estimated standard error and p value derived from a Wald test.

Variable	Estimate	Standard error	p
Intercept	1.077	0.759	0.156
Relative catch (smoothed)	–10.984	1.577	<0.001
Relative catch (smoothed + lag 1)	5.004	1.385	<0.001
Relative catch (smoothed + lag 2)	2.334	1.412	0.098
Relative catch (smoothed + lag 3)	–2.859	0.881	0.001
Catch variability	1.268	0.885	0.152
Maximum recorded length	0.005	0.001	<0.001
Trophic level	–0.765	0.125	<0.001

Table 5. Random effect coefficients estimated as the conditional mode for the restricted maximum likelihood estimate of fixed effects when fitting the logistic regression (eqs. 2–3) to all RAM Legacy Stock Assessment data.

FAO major fishing area code	FAO major fishing area name	Intercept $\widehat{\beta}_f$	Relative catch $\widehat{\beta}_{RC,f}$
21	NW Atlantic	0.9365	2.5607
27	NE Atlantic	2.2192	–6.1232
31	WC Atlantic	0.0841	0.8030
34	EC Atlantic	0.1285	–0.9777
37	Mediterranean and Black Sea	0.1334	–0.6177
41	SW Atlantic	1.0551	1.6371
47	SE Atlantic	–4.1478	7.7769
51	W Indian Ocean	0.0640	–0.2803
57	E Indian Ocean	1.3195	–5.0011
58	Antarctic and southern Indian Ocean	–1.5280	1.6867
67	NE Pacific	–0.6585	1.4610
71	WC Pacific	0.1712	–0.8016
77	EC Pacific	0.7511	0.4432
81	SW Pacific	–1.0291	2.9014
87	SE Pacific	–0.1121	–0.1981
88	Antarctic Pacific	0.0440	–0.2694

validation step by evaluating predictive accuracy on a region-by-region basis or using other covariates (e.g., predictive accuracy varying by fishery size, fish type, etc.). However, we have used a global cross-validation step for simplicity of presentation.

Single-stage models

Other efforts could identify a single-stage model that adequately accounts for model misspecification. Such a model would presumably allow for more natural estimation of a confidence interval of the proportion collapsed by year, which estimation could be improved in the current ACEM. In particular, this would incorporate additional uncertainty about region effects and predictive accuracy for regions with few or no stock assessments.

Additional covariates

Finally, all future efforts to evaluate global fishery status would be aided by linking the databases on fish life history (FishBase), fishery landings (FAO and the Sea Around Us Project), and stock assessments (RAM Legacy Stock Assessment Database) and making the linked data easily available to the public in database form. This would allow the exploration of fine-scale landings statistics and additional life history traits as predictive of stock collapse. Other possible covariates for future investigation include year since a fishery was first developed (Sethi et al. 2010); available fishery-independent survey data; fishing gear types and composition; and model output from existing ecosystem models such as Ecopath with Ecosim (Christensen and Walters 2004) or Atlantis (Fulton et al. 2011).

Formal data collection

The ACEM represents an important bridge between detailed stock assessments with limited coverage and FAO landings data with global coverage. This gap is likely to remain for many years and may never be fully closed because of the great expense involved with collecting, compiling, and analyzing data in formal stock assessments. This gap can also be bridged by using expert opinion on a case-by-case basis as is done in FAO status reports (Food and Agriculture Organization of the United Nations 2010*b*), and model-based methods such as the ACEM are likely to complement expert-opinion methods. We believe that the task of extrapolating global fishery status from stock assessment results can be aided through the expansion of global databases and particularly through the compilation of economic and management information for all stocks globally. We also advocate the development of additional scientific surveys to provide information about abundance trends in data-poor regions of the world. This latter type of time-series data could be incorporated into a model-based estimate of global stock status and would be hugely informative about fishery status globally.

Acknowledgements

Support was provided by the University of Washington, the National Marine Fisheries Service (NMFS) Sea Grant Population Dynamics Fellowship (NA09OAR4170120), an NMFS groundfish project grant to the University of Washington, and National Science Foundation Collaboration on Advanced Multispectral Earth Observation (CAMEO) grants Nos. 1041570 and 1041678. This publication received additional partial funding by the Joint Institute for the Study of the Atmosphere and Ocean (JISA) under National Oceanic and Atmospheric Administration (NOAA) Cooperative Agreement No. 354 NA17RJ1232, Contribution No. 1901. The work was improved by discussions in a University of Washington Think Tank, a CAMEO meeting in Seattle, Washington, 29 November – 1 December 2010 and discussions with A. Punt, C. McGilliard, R. Hilborn, three anonymous reviewers, and the Associate Editor. Finally, to the many research and assessment scientists who contributed to assessments in the RAM Legacy Stock Assessment and the FAO landings databases, thank you!

References

Akaike, H. 1974. New look at statistical-model identification. IEEE Trans. Automat. Contr. **19**(6): 716–723. doi:10.1109/TAC.1974.1100705.

Anticamara, J.A., Watson, R., Gelchu, A., and Pauly, D. 2011. Global fishing effort (1950–2010): Trends, gaps, and implications. Fish. Res. **107**(1–3): 131–136. doi:10.1016/j.fishres.2010.10.016.

Bates, D., and Maechler, M. 2009. lme4: Linear mixed-effects models using S4 classes [online]. Available from http://cran.r-project.org/web/packages/lme4/index.html.

Branch, T.A., Watson, R., Fulton, E.A., Jennings, S., McGilliard, C. R., Pablico, G.T., Ricard, D., and Tracey, S.R. 2010. The trophic fingerprint of marine fisheries. Nature, **468**(7322): 431–435. doi:10.1038/nature09528. PMID:21085178.

Branch, T.A., Jensen, O.P., Ricard, D., Ye, Y., and Hilborn, R. 2011. Contrasting global trends in marine fishery status obtained from catches and from stock assessments. Conserv. Biol. **25**(4): 777–786. doi:10.1111/j.1523-1739.2011.01687.x. PMID:21535149.

Carruthers, T.R., Walters, C.J., and McAllister, M.K. 2012. Evaluating methods that classify fisheries stock status using only fisheries catch data. Fish. Res. **119–120**: 66–79. doi:10.1016/j.fishres.2011.12.011.

Christensen, V., and Walters, C.J. 2004. Ecopath with Ecosim: methods, capabilities and limitations. Ecol. Model. **172**(2–4): 109–139. doi:10.1016/j.ecolmodel.2003.09.003.

Costello, C., Gaines, S.D., and Lynham, J. 2008. Can catch shares prevent fisheries collapse? Science, **321**(5896): 1678–1681. doi:10.1126/science.1159478. PMID:18801999.

Daan, N., Gislason, H., Pope, J.G., and Rice, J.C. 2011. Apocalypse in world fisheries? The reports of their death are greatly exaggerated. ICES J. Mar. Sci. **68**(7): 1375–1378. doi:10.1093/icesjms/fsr069.

de Mutsert, K., Cowan, J.H., Jr, Essington, T.E., and Hilborn, R. 2008. Reanalyses of Gulf of Mexico fisheries data: landings can be misleading in assessments of fisheries and fisheries ecosystems. Proc. Natl. Acad. Sci. U.S.A. **105**(7): 2740–2744. doi:10.1073/pnas.0704354105. PMID:18287085.

Food and Agriculture Organization of the United Nations. 2010*a*. Capture production 1950–2009. FISHSTAT Plus — Universal software for fishery statistical time series [online or CD-ROM]. Food and Agriculture Organization of the United Nations, Statistics and Information Service of the Fisheries and Aquaculture Department, Rome, Italy. Available from http://www.fao.org/fishery/statistics/software/fishstat/en.

Food and Agriculture Organization of the United Nations. 2010*b*. The state of the world fisheries and aquaculture 2010. Food and Agriculture Organization of the United Nations, Fisheries and Aquaculture Department, Rome, Italy.

Froese, R., and Pauly, D. 2010. FishBase [online]. Available from http://fishbase.org/.

Fulton, E.A., Link, J.S., Kaplan, I.C., Savina-Rolland, M., Johnson, P., Ainsworth, C., Horne, P., Gorton, R., Gamble, R.J., Smith, A.D.M., and Smith, D.C. 2011. Lessons in modelling and management of marine ecosystems: the Atlantis experience. Fish Fish. **12**(2): 171–188.

Greenberg, P. 2010. Time for a sea change. Natl. Geogr. Mag. **218**: 70–89.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. Springer-Verlag Inc., New York.

Hilborn, R. 2011. Let us eat fish [online]. New York Times. 14 April 2011. Available from http://www.nytimes.com/2011/04/15/opinion/15hilborn.html?_r=1.

Hosmer, D.W., and Lemeshow, S. 2000. Applied logistic regression. Wiley-Interscience Publication, Danvers, Mass.

Hutchings, J.A., Minto, C., Ricard, D., Baum, J.K., and Jensen, O.P. 2010. Trends in the abundance of marine fishes. Can. J. Fish. Aquat. Sci. **67**(8): 1205–1210. doi:10.1139/F10-081.

Maunder, M.N., and Punt, A.E. 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. **70**(2–3): 141–159. doi:10.1016/j.fishres.2004.08.002.

Mullon, C., Fréon, P., and Cury, P. 2005. The dynamics of collapse in world fisheries. Fish Fish. **6**(2): 111–120.

Pauly, D. 2009. Aquacalypse now: the end of fish. New Repub. **240**(18): 24–27.

Pauly, D. 2010. 5 Easy pieces: the impact of fisheries on marine ecosystems. Island Press, Washington, D.C.

Pauly, D., and Froese, R. 2012. Comments on FAO's State of Fisheries and Aquaculture, or "SOFIA 2010." Mar. Policy, **36**(3): 746–752. doi:10.1016/j.marpol.2011.10.021.

Pinsky, M.L., Jensen, O.P., Ricard, D., and Palumbi, S.R. 2011. Unexpected patterns of fisheries collapse in the world's oceans. Proc. Natl. Acad. Sci. U.S.A. **108**(20): 8317–8322. doi:10.1073/pnas.1015313108. PMID:21536889.

R Development Core Team. 2009. R: a language and environment for statistical computing [online]. Available from http://cran.r-project.org/.

Ricard, D., Minto, C., Jensen, O.P., and Baum, J.K. Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. Fish Fish. [In press.] doi:10.1111/j.1467-2979.2011.00435.x.

Sethi, S.A., Branch, T.A., and Watson, R. 2010. Global fishery development patterns are driven by profit but not trophic level. Proc. Natl. Acad. Sci. U.S.A. **107**(27): 12163–12167. doi:10.1073/pnas.1003236107.

Walters, C. 2003. Folly and fantasy in the analysis of spatial catch rate data. Can. J. Fish. Aquat. Sci. **60**(12): 1433–1436. doi:10.1139/f03-152.

Worm, B., Barbier, E.B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, B.S., Jackson, J.B.C., Lotze, H.K., Micheli, F., Palumbi, S.R., Sala, E., Selkoe, K.A., Stachowicz, J.J., and Watson, R. 2006. Impacts of biodiversity loss on ocean ecosystem services. Science, **314**(5800): 787–790. doi:10.1126/science.1132294. PMID:17082450.

Worm, B., Hilborn, R., Baum, J.K., Branch, T.A., Collie, J.S., Costello, C., Fogarty, M.J., Fulton, E.A., Hutchings, J.A., Jennings, S., Jensen, O.P., Lotze, H.K., Mace, P.M., McClanahan, T.R., Minto, C., Palumbi, S.R., Parma, A.M., Ricard, D., Rosenberg, A.A., Watson, R., and Zeller, D. 2009. Rebuilding global fisheries. Science, **325**(5940): 578–585. doi:10.1126/science.1173146. PMID:19644114.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., and Smith, G.M. 2009. Mixed effects models and extensions in ecology with R. Springer, New York.