JARA: 'Just Another Red-List Assessment'

Henning Winker[1,2] *& Richard B. Sherley*[3]

[1] Department of Agriculture, Forestry and Fisheries (DAFF), Roggebaai, 8012 Cape Town, South Africa
[2] Centre for Statistics in Ecology, Environment and Conservation (SEEC), Department of Statistical Sciences, University of Cape Town, South Africa.
[3] Environment and Sustainability Institute, College of Life and Environmental Sciences, University of Exeter, Penryn Campus, Cornwall, TR10 9FE, UK

Correspondance: Henning Winker. E-mail: henning.winker@gmail.com; Richard B. Sherley. Email: r.sherley@exeter.ac.uk.

Abstract

Identifying species at risk of extinction is necessary to prioritise conservation efforts. The International Union for Conservation of Nature's (IUCN) Red List of Threatened Species is the global standard for quantifying extinction risk, with many species categorised on the basis of a reduction in population size. We introduce the Bayesian state-space framework 'JARA' (Just Another Red-List Assessment). Designed as decision-support tool, JARA allows both process error and uncertainty to be incorporated into IUCN Red List assessments under criterion A. JARA outputs easy to interpret graphics showing the posterior probability of the population decline displayed against the IUCN Red List categories, and assigns each category an associated probability given process and observation uncertainty. JARA is designed to be easy to use, rapid and widely applicable, so conservation practitioners can apply it to their own count or relative abundance data. We illustrate JARA using three real-world examples: (1) relative abundance indices for two elasmobranchs, Yellowspotted Skate *Leucoraja wallacei* and Whitespot Smoothhound *Mustelus palumbes*; (2) a comparison of standardized abundance indices for Atlantic Blue Marlin *Makaira nigricans* and (3) absolute abundance data for Cape Gannets *Morus capensis*. Finally, using a simulation experiment, we demonstrate how JARA provides greater accuracy than two approaches commonly used to assigning a Red List Status under criterion A. Tools like JARA can help further standardise Red List evaluations, increasing objectivity and lowering the risk of misclassification. The consequences for global conservation efforts could be substantial.

Introduction

Quantifying trends in population abundance is central to ecological research and to conservation biology in particular, where understanding extinction risk is necessary to prioritise effort in the

face of ever-increasing biodiversity loss (Butchart et al. 2004; Hoffmann et al. 2008). Although a number classification protocols exist to assess a species' extinction risk (Regan et al. 2013), the International Union for the Conservation of Nature's (IUCN) Red List of Threatened Species is viewed widely as the global standard (Hoffmann et al. 2008; Mace et al. 2008).

To list a species in a threatened category (vulnerable [VU], endangered [EN] or critically endangered [CR]) on the Red List, assessors can consider the risks associated with both the small-population paradigm and the declining population paradigm (Caughley 1994) under five assessment criteria (A to E). However, quantitative information on species abundance – as a direct measure of the status of a population – is often preferred over less direct measures of extinction risk, like changes in habitat extent or quality (Mace et al. 2008; Wilson et al. 2011). Moreover, the IUCN guidelines allow for assignment to a category (ranging from Least Concern [LC] to Extinct) based only on the criterion that produces the highest estimated risk (Mace et al. 2008). Therefore, species are often listed only on the basis of a reduction in population size (Criterion A); as of today > 5,000 species are classified as threatened on this basis alone (Rueda-Cediel et al. 2018). Criterion A is thus considered to be the most widely used stand-alone criterion for assigning a Red List status to a wide range of animal taxa, including mammals, birds, reptiles, fishes and insects (Butchart et al. 2004; Dulvy et al. 2006; Mace et al. 2008; Seminoff & Shanker 2008; Fox et al. 2018).

Although the IUCN Red List provides a set of unified quantitative decision rules for assigning threatened categories based on population decline thresholds, data quality and analytical approaches can differ vastly (Mace et al. 2008; Wilson et al. 2011). Estimating the decline between just two points in time remains widely applied as a rapid assessment approach (Wilson et al. 2011; IUCN 2017; Lee et al. 2019). Even when model-predicted abundance indices or population trajectories with associated uncertainty estimates exist, it appears common practice to simply extract the two points that come closest to the desired IUCN assessment time frame of 10 years or (if longer) three generation lengths and then calculate the decline as the ratio between these two points. If the observation period is shorter than three generation lengths, a simple extrapolation formula may be applied (IUCN 2017; Lee et al. 2019). Fitting linear or log-linear regression models to time series of abundance observations is a second widely-applied approach (Dulvy et

al. 2006; Wilson et al. 2011; IUCN 2017; Fox et al. 2018). An advantage of using a regression is that estimates of the significance of the estimated slope, parameter uncertainty, and goodness-of-fit can be obtained, which is not possible with only two points (Wilson et al. 2011). However, regression implicitly assumes that the "true" population follows a constant deterministic trend, with all deviations from this trend attributable to statistically-independent observation errors (Hilborn & Mangel 1997). This assumption does not agree well with our understanding of the real world. For example, a population diminished due to a severe drought year or a spike in poaching will start from a lower state than would be expected from an average long-term trend. In reality, animal population trajectories can diverge substantially from this deterministic expectation due to persistent variations in environmental regimes or human impacts (Connors et al. 2014; D'Eon-Eggertson et al. 2015; Keith et al. 2015).

In contrast to deterministic regression models, state-space models provide a general framework for analysing dynamical systems that considers process (year-to-year variation) and observation (or reporting) error simultaneously (de Valpine & Hastings 2002; Kéry & Schaub 2012). State-space formulations add more biological realism by assuming a Markovian process, such that the population size in the next time step is conditioned on its current state, but independent of past and future states (Kéry & Schaub 2012). From a statistical perspective, the observations are independent given the unobservable (latent) states of population. This can help in preventing violation of independence in observation errors, which often manifest themselves as serial residual correlations when fitting simple regressions to time-series data. Given these desirable properties, it is not surprising that there has been a rapid uptake of state-space approaches for modelling population dynamics (de Valpine & Hastings 2002; Buckland et al. 2004; Thorson & Minto 2015), including more recent applications of quantifying population decline (Wilson et al. 2011; Bauer et al. 2015; Simmons et al. 2015; Boyd et al. 2017). State-space models have also been specifically applied in simulation studies that highlight the need for a more rigorous quantification of process and observation uncertainties in the context of threat category (mis)classification (Wilson et al. 2011; Regan et al. 2013; Connors et al. 2014).

Bayesian implementations of state-space models offer a powerful framework to improve the characterization and communication of uncertainty during IUCN Red List assessments. The

posterior probabilities provide an intuitive and transparent way to express uncertainty about population declines to conservation practitioners (Bauer *et al.*, 2015; Sherley *et al.*, 2018), which can be translated directly into probabilistic statements about a population falling into a threatened category (Bauer et al. 2015; Boyd et al. 2017). However, developing customized Bayesian state-space models can be technically demanding and time consuming, especially when dealing with often case-specific, 'messy' abundance data that are subject to missing values, irregular spacing or multiple indices that are measured at different scales. These issues may therefore dissuade some conservation practitioners and hamper broader applications.

To address this, we have developed the Bayesian state-space framework 'JARA' (Just Another Red-List Assessment). Designed as an easy to use, rapid and widely applicable decision-support tool, JARA allows both process error and uncertainty to be incorporated into Red List assessments. A key output of JARA is an easy to interpret graphic in which the probability distribution of the population decline is displayed against the IUCN Red List categories, and where each category is assigned a probability given process and observation uncertainty. To ensure a high degree of transparency and reproducibility, we provide fully commented R code on the global open-source platform GitHub (https://github.com/henning-winker/JARA), so that JARA can be modified and applied by conservation practitioners to their own count or relative abundance data. On GitHub, we provide a number of worked examples. Here, we illustrate the main features of JARA using three real-world examples and a simulation experiment to examine how accurately JARA classifies complex population declines under criterion A, relative to the regression and "two-points" approaches discussed above.

Material and Methods

JARA is a generalized Bayesian state-space decision-support tool for trend analysis of abundance indices with direct applications to IUCN Red List assessments. The name 'Just Another Red-List Assessment' acknowledges JAGS (Just Another Gibbs Sampler, Plummer, 2003), the software used to run the Bayesian state-space model application. The name reference, together with user-friendly R interface and modulated coding structure follows the example of the new open source fisheries stock assessment software 'Just Another Bayesian Biomass Assessment' (JABBA; Winker et al. 2018). JARA enables analysis of one or multiple abundance indices simultaneously,

where each index can contain missing years and span different periods. JARA provides the option for fitting relative abundance indices to estimate a mean trend or absolute abundance indices (from e.g. different subpopulations) to produce a summed population trend for the total population.

State-Space Model Formulation

A central assumption of the state-space approach is that the abundance (I_t) trend follows a Markovian process, such that I_t in year *t* will be conditioned on I_{t-1} in the previous year. For generality it is assumed that the underlying population trend follows a conventional exponential growth model $I_{t+1} = I_t\lambda_t$ (e.g. Kéry & Schaub 2012), where λ_t is the growth rate in year *t*. The growth rate λ_t can vary annually to accommodate fluctuations in reproductive success and survival as a result of environmental conditions, anthropogenic activity or other latent (unobservable) impacts. State-space models are hierarchical models that explicitly decompose an observed time-series into a process variation and an observation error component (Simmons et al. 2015). On the log scale, the process equation becomes $\mu_{t+1} = \mu_t + r_t$, where $\mu_t = \log(I_t)$ and $r_t = \log(\lambda_t)$ is the year-to-year rate of change, with variations in log-growth rates following a random normal walk $r_t \sim Normal(\bar{r}_p, \sigma_\eta^2)$, given the estimable process error variance σ_η^2 and the estimable mean population rate of change $\bar{r}_p$ (i.e. the underlying trend). The corresponding observation equation is of the form $\log(y_t) = \mu_t + \epsilon_t$, where y_t denotes the abundance value for year *t*, ϵ_t is observation residual for year *t*, which is assumed to be normally distributed on log-scale $\varepsilon_t \sim Normal(0, \sigma_\varepsilon^2)$ as a function of the observation variance σ_ε^2.

The model notation for relative abundance indices builds on the approach in JABBA for averaging relative abundance indices (Winker et al. 2018) and assumes that the mean underlying abundance trend is an unobservable state variable. The corresponding observation equation is then modified, such that $\log(y_{t,i}) = \mu_t + \log(q_i) + \epsilon_{t,i}$, where $y_{t,i}$ is the relative abundance value for year *t* and index *i*, μ_t is the natural logarithm of the mean abundance trend, $\varepsilon_{t,i}$ is the lognormal observation error term for index *i* and year *t*, and q_i is a scaling parameter for index *i*. The abundance index with the chronologically oldest record is taken as a reference index by fixing $q_1 = 1$ and the other indices are scaled to this reference index, respectively, with $q_{2,\ldots,n}$ being estimable model

parameters. The estimated posterior of the population trend for year t is then given by $I_{p,t} = \exp(\mu_t)$.

To estimate a total abundance trajectory for the "global" population from multiple absolute abundance indices, we assume that each absolute abundance index represents a 'subpopulation' that may increase or decline independently from other subpopulations. The process equation is therefore modified to $\mu_{t,i} = \mu_{t,i} + r_{t,i}$, where $r_{t,i} = \log(\lambda_{t,i})$ is the year-to-year rate of change specific to index *i* that is assumed to vary around $\bar{r}_i$ to represent the underlying mean rate of change for the subpopulation (instead of a global population mean $\bar{r}_p$), but with a process variance σ_η^2 that is common to all subpopulations: $r_{t,i} \sim Normal(\bar{r}_i, \sigma_\eta^2)$. The corresponding observation equation is adjusted to $\log(y_{t,i}) = \mu_{t,i} + \epsilon_{t,i}$, so that abundance trend $\mu_{t,i}$ and the error term $\varepsilon_{t,i} \sim Normal(0, \sigma_{\varepsilon_i}^2)$ now become the specific to subpopulation *i*. The estimated posterior of the global population trajectory $I_{p,t}$ for year t is then estimated from the sum of all individual subpopulation trajectory posteriors, such that $I_{p,t} = \sum_i \exp(\mu_{t,i})$.

Bayesian Framework

JARA is run from the statistical environment R (R Core Team 2018) and executed in JAGS (Plummer 2003), using a wrapper function from the R library 'r2jags' (Su & Yajima 2012). The Bayesian posterior distributions are estimated by means of Markov Chain Monte Carlo (MCMC) simulation. In JAGS, all estimable hyper-parameters have to be assigned to a prior distribution. JARA uses vague (uninformative) prior distributions throughout, so all inferences are drawn from the information in the data. The estimation of annual growth rate deviates r_t, is implemented through hierarchical priors (Jiao et al. 2009), where r_t is informed by the population mean $\bar{r}_p$ for relative abundance indices and $r_{t,i}$ is informed by $\bar{r}_i$ for absolute abundance indices. Normal prior distributions of $prior \sim Normal(0,1000)$ are assumed for both $\bar{r}_p$ or $\bar{r}_i$. The initial population size in the first year $I_{t=1,i}$ is drawn in log-space from a 'flat' normal distribution with the mean equal to the log of the first available count $y_{t=1,i}$ and a standard deviation of 1000. Priors for the process variance can be either fixed or estimated. If estimated (default), the process variance prior is implemented via a vague inverse-gamma distribution by setting both scaling parameters to 0.001

(Chaloupka & Balazs 2007; Brodziak & Ishimura 2012; Carvalho et al. 2014), which yields an approximately uniform prior on the log scale.

The total observation variance σ_{ε}^2 can be separated into three components: (1) externally derived standard errors $\sigma_{SE_{y,i}}^2$ for each abundance index *i*, (2) a fixed input variance σ_{fix}^2 and (3) estimable variance σ_{est}^2, where the prior for σ_{est}^2 assumes an uninformative inverse-gamma distribution with both scaling parameters set to 0.001. All three variance components are additive in their squared form (Francis et al. 2003), so the total observation variance for abundance index *i* and year *y* is: $\sigma_{\varepsilon_{y,i}}^2 = \sigma_{SE_{y,i}}^2 + \sigma_{fix}^2 + \sigma_{est}^2$. Each variance component can be individually switched on or off by the user (Winker et al. 2018). Adding a fixed observation error to externally estimated standard errors is common practice to account for additional sampling error associated with abundance indices and also informs the process variance as a portion of total variance is assigned *a priori* to observation variance (Winker et al. 2018).

Estimating probabilities of population decline

A posterior probability for the percentage change (*C%*) associated with each abundance index can be conveniently calculated from the posteriors of $\hat{I}_{p,y}$, the model predicted population trajectory. If $\hat{I}_{p,y}$ represents a longer time span than the assumed three generation lengths (GL), *C%* is automatically calculated as the difference between a three-year average around the final observed data point T, and a three-year average around the year corresponding to $T - (3 \times GL)$. The year $T + 1$ is always projected to obtain a three-year average around T to reduce the influence of short-term fluctuations (Froese et al. 2017). When the span of $\hat{I}_t$ is < 3 × GL, JARA projects forward, by passing the number of desired future years without observations to the state-space model until $\hat{I}_t > (3 \times GL) + 2$. The projections for a single population or multiple subpopulations *i* are based on all possible posterior realizations of $r_{t,i}$ across all *T* years in the observed time series: $\tilde{r}_i = \frac{1}{T}\sum_{t=1}^{T} r_{t,i}$.

Model Diagnostics

To evaluate the model fit, JARA provides the user with four plots. The first illustrates the unscaled input data and uncertainty estimates around each observation in the form 95% Confidence Intervals

(Figure 1a-b). The second shows the observed and predicted abundance values for each time series together with the 95% posterior predictive credibility intervals (Figure 1c-d). The third shows individual fits on the log-scale, as well as the 95% Bayesian credible intervals (CI) derived from the observation variance $\sigma^2_{\varepsilon_{y,i}}$ (Figure 2). The fourth is a residual plot (Figure 1e-f) to illustrate potential data conflict when fitting multiple time series (Winker et al. 2018). This plot includes: (1) colour-coded lognormal residuals of observed versus predicted abundance indices *i*, (2) boxplots indicating the median and quantiles of all residuals available for each year; the area of each box indicates the strength of the discrepancy between the abundance indices (larger box means higher degree of conflicting information), and (3) a loess smoother through all residuals which highlights systematically auto-correlated residual patterns (Figure 1e-f). In addition, the Root-Mean-Squared-Error (RMSE) and the deviance information criterion (DIC) are provided for the comparison of goodness-of-fit and model selection purposes respectively. Convergence of the MCMC chains is diagnosed using the 'coda' package (Plummer et al. 2006), adopting minimal thresholds of $p = 0.05$ for Heidelberger and Welch (Heidelberger & Welch 1992) and Geweke's (Geweke 1992) diagnostics called from R.

JARA decision support plots

To facilitate decision-making, JARA routinely produces four key plots (Figures 3–5) for each input dataset showing: (1) the combined state-space model fit and 95% CI to abundance time-series where multiple indices relative abundance (e.g. Figure 4a) and stock assessment results (Figure 4b) are used, or the individual state-space models fits to each count dataset for absolute abundance indices (e.g. individual colony counts for Cape Gannet; Figure 5a); (2) the overall observed and projected (±95% CI) population trajectory over three GL (e.g. Figure 3a and b); (3) the median and posterior probabilities for the percentage annual population change calculated from all the observed data, and from each of the most recent 1 GL, 2 GL, and 3 GL (depending on the length of the observed time-series), shown relative to a stable population ($\%C = 0$) (e.g. Figure 5c); and (4) how the posterior distribution for the percentage change in abundance ($\%C$) over 3 GL aligns against the thresholds for the Red List categories (Least Concern LC, Near Threatened NT, Vulnerable VU, Endangered EN or Critically Endangered CR) under criteria A2–A4 (e.g. Figure 4e and f) or under A1. In addition, JARA can be used to undertake a retrospective analysis through the sequential removal of terminal years and subsequent forward projections to attain 3 GL

("retrospective peel"). This enables to user to identify points in time where the *%C* crosses over into different Red List categories (Figure 5f).

Simulation experiment

We conducted a simulation experiment to compare the performance of JARA against two conventional approaches suggested in the IUCN Red List guidelines: (1) the 'two-point' approach that calculates %*C* between two years with observations that are 3 × GL apart and (2) regression analysis that extrapolates %C from the estimated slope over a given assessment horizon of 3 × GL (IUCN 2017). We first used an operating model (OM) to generate a 'true' population trend and an 'observed' abundance index (Figure S1). The observed abundance index is then passed to the three estimation models (EMs: 'Two-Points', 'Regression', 'JARA') to determine the likely threat category (Figure 6).

We modelled the 'true' population numbers using a log-linear Markovian process, such that $\log(N_t) = \log(N_{t-1}) + r_t$, where r_t determines the annual rate of change in the population numbers N_t. We assumed that r_t can be decomposed into three processes: (i) an underlying mean population trend $\bar{r}$, (ii) stochastic annual variation in the population υ_t and (iii) annual variation in anthropogenic impact τ_t, such that $r_t = \bar{r} + \upsilon_t - \tau_t$ (Fig. S1). The process error deviations associated with υ_t and τ_t were assumed to be normally distributed and temporally autocorrelated (Connors et al. 2014; Keith et al. 2015) by implementing first-order autoregressive functions (AR1) of the form: $\upsilon_t = \upsilon_{t-1}\phi_\upsilon + \delta_t\sqrt{1-\phi_\upsilon^2}$ and $\tau_t = \upsilon_{t-1}\phi_\tau + \gamma_t\sqrt{1-\phi_\tau^2}$ where ϕ_υ and ϕ_t are the AR1 autocorrelation coefficients and $\delta_t \sim Normal(0, \sigma_\delta^2)$ and $\gamma_t \sim Normal(0, \sigma_\gamma^2)$ are normally distributed random variables representing uncorrelated process errors with variances σ_δ^2 and σ_γ^2, respectively (Thorson et al. 2014). The observed abundance index I_y is then generated as a function of the 'true' population numbers and a lognormal observation error of the form: $I_y = N_t e^{\varepsilon_t}$, where $\varepsilon_t \sim Normal(0, \sigma_\varepsilon^2)$ and σ_ε^2 is the an observation variance σ_ε^2. A generic R function of the population simulator is provided in the Supplementary Information.

We simulated 200 random sets of population time series of 35 years with an assumed assessment horizon of 30 years corresponding to a GL of 10 years (Figure 6). The initial population size was

$N_1 = 10{,}000$ and random deviates of $\bar{r}$, describing the underlying trend, were randomly generated using a uniform distribution between −0.05 and 0 to cover a wide range of population trends and associated *%C* (Figure S1-2). We set the standard deviation for the stochastic variation to a moderate value of $\sigma_\upsilon = 0.1$ with a process error autocorrelation of $\phi_\upsilon = 0.3$. This choice of values was intended to describe a population where annual reproductive success is fairly variable and autocorrelation arises from individual age-classes of varying strengths growing through the population. By comparison, the variation in anthropogenic impact was assumed to be less variable $\sigma_\upsilon = 0.05$ and more persistent over longer, regime-like periods by assuming a fairly large autocorrelation coefficient of $\phi_\tau = 0.7$. The observation error for the observed abundance was set to $\sigma_\varepsilon = 0.15$ (i.e. CV ~ 15%). The resulting simulations produced a wide range of plausible population trajectories (Figure 6).

We evaluated model performances in terms of the accuracy of the estimated 'true' %*C* over 3 GL and the proportion of correctly classified threat categories according to the IUCN Red List A2 decline criterion. Accordingly, the 'true' %*C* over 30 years (3 *GL*) was calculated as $\%C = (N_{35}/N_6 - 1) \times 100$. For the 'Two-Point' EM we calculated $\%\hat{C} = (I_{35}/I_6 - 1) \times 100$. For the 'Regression' EM we closely followed the IUCN guidelines by: (i) fitting a log-linear regression of the form $\log(I_y) = \alpha + \beta y$ to all observations I_{1-35}, (ii) predicting the expected values $\hat{I}_{1-35}$, (iii) and then extrapolating the expected $\%\hat{C}$ over 3 GL as: $\%\hat{C} = \left(\hat{I}_{35}/\hat{I}_1\right)^{(3GL/n_y)} - 1) \times 100$ where n_y is the number of years. As described above (*Estimating probabilities of population change*), for the JARA EM the $\%\hat{C}$ was calculated as the median of the posterior %*C*. Accuracy and bias were calculated using Median Absolute Error (MAE) and Median Error (ME), where the error is the difference between the estimated $\%\hat{C}$ and the 'true' $\%C$.

Results

<u>*Case study applications*</u>

We illustrate applications of JARA using the following worked examples for different abundance data types: (1) multiple relative abundance indices (1991–2017) from South African scientific demersal trawl surveys for two elasmobranchs, Yellowspotted Skate *Leucoraja wallacei* and Whitespot Smoothhound *Mustelus palumbes*; (2) a comparison of model-predicted biomass

trajectories from the 2018 Atlantic Blue Marlin *Makaira nigricans* stock assessments and standardized catch-per-unit-effort (CPUE) abundance indices from various fishing fleet, which were used as input into the assessment model and (3) census data for a large seabird, the Cape Gannet *Morus capensis*, in the form of total breeding size estimates for its six global breeding localities in Namibia and South Africa (1979–2017). All worked examples and input data are available on GitHub (https://github.com/henning-winker/JARA).

(1) White-spotted smoothhound and Yellowspotted skate are both endemic to southern Africa and had been previously listed as Data Deficient by the IUCN (Smale 2009; Smale 2006). During the 2018 Southern African IUCN Red List workshop, hosted by the IUCN Shark Specialist Group (SSG), both species were reassessed using JARA as a decision support tool. Based on available life-history information, the GLs were empirically estimated as 14 years for Whitespot Smoothhound and 12 years for Yellowspotted Skate. Long-term abundance indices were obtained from analysis of demersal research trawl survey data that were conducted with alternating trawl gear configurations (old/new gear) during summer along South Africa's west coast (1991–2017) and during autumn and spring along the south coast (1991–2016) by the Department of Agriculture, Forestry and Fisheries. Yellowspotted skate was only encountered during South Coast Surveys. Annual density estimates (kg per nm^2 area swept) and associated standard errors on log-scale were estimated using a geostatistical delta-GLMM (Thorson et al. 2015, 2016). We only considered the period from 1991 onwards due to evident learning effects for identifying Chondrichthyes during initial survey years. The estimated abundance indices were separately derived for each gear type, survey (coast/season). The input data are characteristic of multiple indices with irregular spacing due to missing survey years (Figure 1 and 2), and are estimated at varying spatial scales. They are therefore treated as relative abundance indices.

JARA combined six abundance indices for Whitespot Smoothhound (Figure 1a) and four abundance indices for Yellowspotted Skate (Figure 1b). Both sets of abundance indices spanned ~2 GL, so JARA automatically projected over ~ 1 GL to provide $\%C$ estimates of +20.2% with 95% of the posterior falling in LC for Whitespot Smoothhound and −47.1% for Yellowspotted Skate, with EN the best-supported category (Figure 3). The plots of the percentage annual population change (Figure 3c and d) nicely illustrate the contrasting situations for these two

species, with Whitespot Smoothhound showing an improving population growth rate in the most recent 1 GL (relative to the whole dataset, Figure 3c), while for Yellowspotted Skate the decline appears to have worsened over the last 1 GL (Figure 3d).

(2) The blue marlin, one of the most iconic gamefishes, is a highly migratory billfish species distributed circumglobally in mostly tropical waters. The last IUCN assessment in 2010 classified the species globally as VU (Collete et al. 2011). The Atlantic population was assessed by fitting linear regressions to biomass trend estimates from stock assessments undertaken by the International Commission for the Conservation of Atlantic Tunas (ICCAT; 2011). GL was assumed to be up to 6 years, corresponding to an IUCN trend assessment horizon of 18 years (3 x GL). Here, we use the 2018 ICCAT stock assessment of Atlantic Blue Marlin to illustrate how JARA can objectively convert such stock assessment outputs into corresponding IUCN Red List decline thresholds and threat categories recommendations. The assessment output used as input for JARA is the median of the estimated biomass for the period 1950–2016 across all accepted candidate models (Figure 4b). Uncertainty was represented by annual standard errors (SEs) of the log-biomass estimates (ICCAT 2018). We then compared the JARA results with a complementary JARA analysis of multiple CPUE indices that served as input data into the 2018 stock assessment. This CPUE dataset comprised standardized CPUE indices (+SEs) from ten different fishing fleets that covered different periods between 1959 and 2016 (Figure 4a).

JARA estimated similar median $\%C$ for the Atlantic Blue Marlin whether using the multiple CPUE indices (−29.3%) or the stock biomass (−21.8%) as input data. In both cases, the best-supported category was LC, but uncertainty was (unsurprisingly) lower when using the biomass estimates from the stock assessment (80.7% of the posterior in LC) versus the much more variable and partially conflicting CPUE indices (51.4% in LC; Figure 4e and f). Although either input series would likely lead to the same outcome (a classification of LC) if JARA were used a decision-support tool to assess Atlantic Blue Marlin, the analysis of the stock assessment data would result in greater confidence in the assessment's conclusion.

(3) The Cape Gannet is a large seabird endemic to southern Africa, where it breeds during the austral spring and summer at six islands. Aerial photographs and on the ground measurements of

nesting density at all six breeding colonies have been used by the Department of Environmental Affairs (South Africa) and the Ministry of Fisheries and Marine Resources (Namibia) to estimate the total breeding population since the summer of 1956/57 (Crawford et al. 2007). Counts were sporadic until 1978/79 but undertaken approximately annually thereafter (Figure 5). Cape Gannets generally have high adult survival, strong breeding site fidelity (Distiller et al. 2012) and low reproductive success (≤ 0.82 chicks fledged per pair annually; Crawford et al. 2019). Accordingly, their GL is estimated as 20.2 years (BirdLife-International 2018) and we use annual count data from 1956/57 to 2016/2017, which spans 3 GL or the recommended Red List assessment period. Of a possible 366 annual counts between 1956/57 to 2016/2017, 173 were not made for various reasons. Thus, the Cape Gannet dataset is chosen an example of multiple absolute abundance indices (with missing data) that in sum represents the trend of the global population.

The output for Cape Gannet nicely exemplifies the decision-support element of JARA. Based on only the median $\%C$ (−51.3%), the species meets the criteria for classification as EN (Figure 5d). However, the uncertainty spans LC to EN; assessors may, therefore, also want to consider that while 55.6% of the posterior probability distribution falls into EN, 41.8% also falls into VU (Figure 5d). Moreover, the percentage annual population change plots show that the decline has slowed and then accelerated over the last 2 GL (Figure 5c), while the retrospective plots show that the threshold for EN was first exceeded in 2013, but the situation has not worsened markedly to 2016. Assessors could then make an informed decision on whether to list the Cape Gannet as VU or EN, depending on how risk prone or adverse they wanted to be in their assessment and the presence of other mitigating factors.

Simulation experiment results

Median Absolute Errors (MAEs) between the estimated and 'true' percentage change over 3 generation lengths (GL) using JARA as the EM was 3.63%. The Regression EM (7.87%) and Two-Point EM (10.96%) had MAEs >2 and >3 times larger, respectively, showing a distinct decrease in the accuracy of the estimated $\%C$ in comparison to the JARA EM (Figure 7). In addition, although all three EMs showed a tendency towards a positive bias (Median Errors > 0), this bias was ~3.7 times larger for the regression EM (0.71%) and ~12 times larger for the Two-Point EM (2.26%) than for JARA (0.19%). Moreover, the median $\%C$ from JARA correctly identified the

'true' IUCN threat category in 89.5% of cases; > 10% more than the regression EM (which was based on the current IUCN guidelines) and >20% more than the Two-Point EM (Figure 7).

Figures

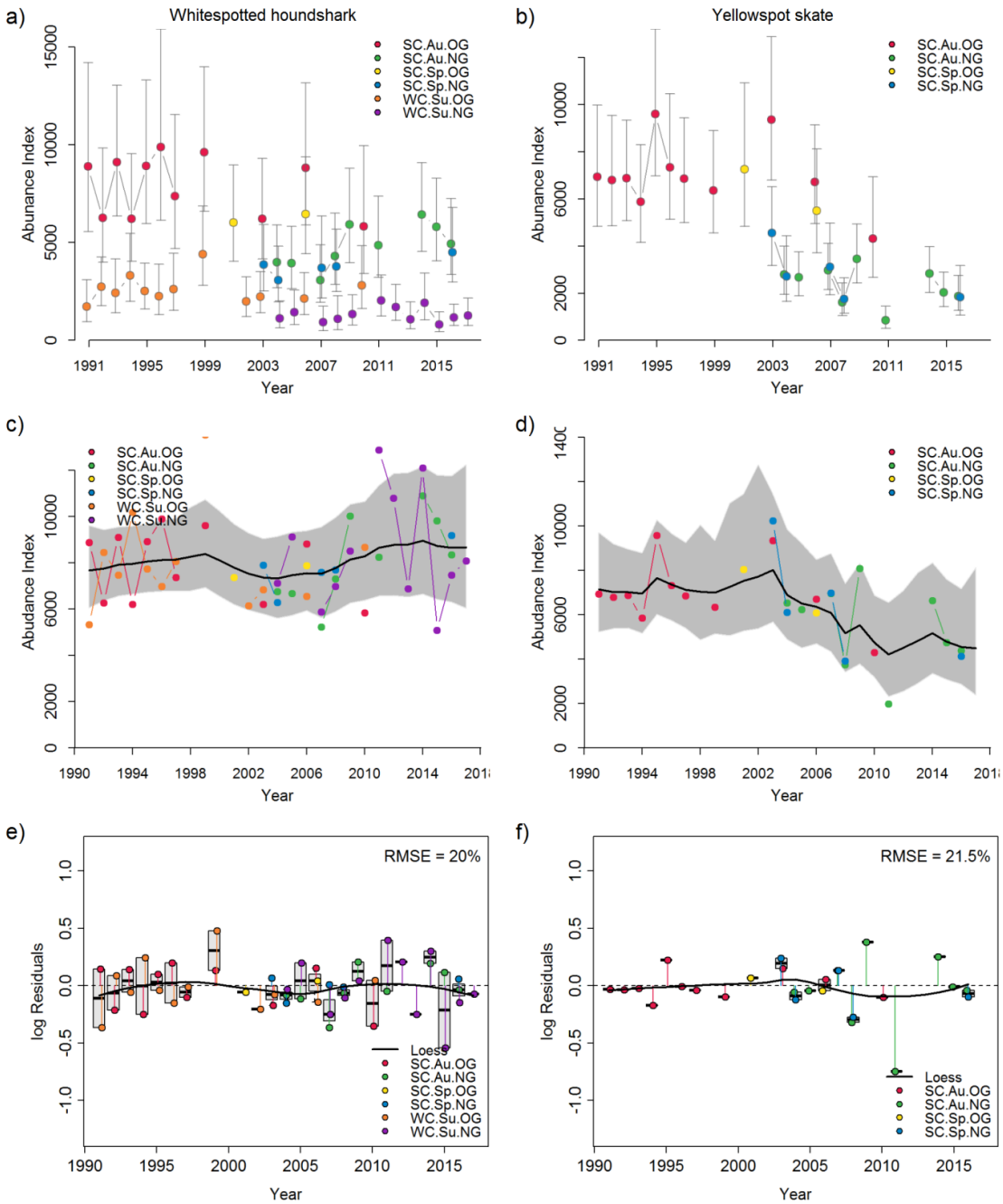

Figure 1: Input datasets of standardized relative abundance indices (± 95% CIs) from South African trawl surveys for two chondrichthyes species, (a) Whitespot Smoothhound *Mustelus palumbes* and (b) Yellowspotted Skate *Leucoraja wallacei* and Model diagnostics plots produced by JARA showing (c) – (d) scaled observed relative abundance (coloured points) for each catch per unit effort (CPUE) timeseries with model fits (predicted relative abundance, black line) and 95% posterior predictive credibility intervals (grey polygons) and (e) – (f) colour-coded lognormal residuals of observed versus predicted abundance indices (coloured points), boxplots (grey bars) indicating the median and quantiles of all residuals available for any given year, and a loess smoother (black line) through all residuals to highlight any systematically auto-correlated residual patterns.

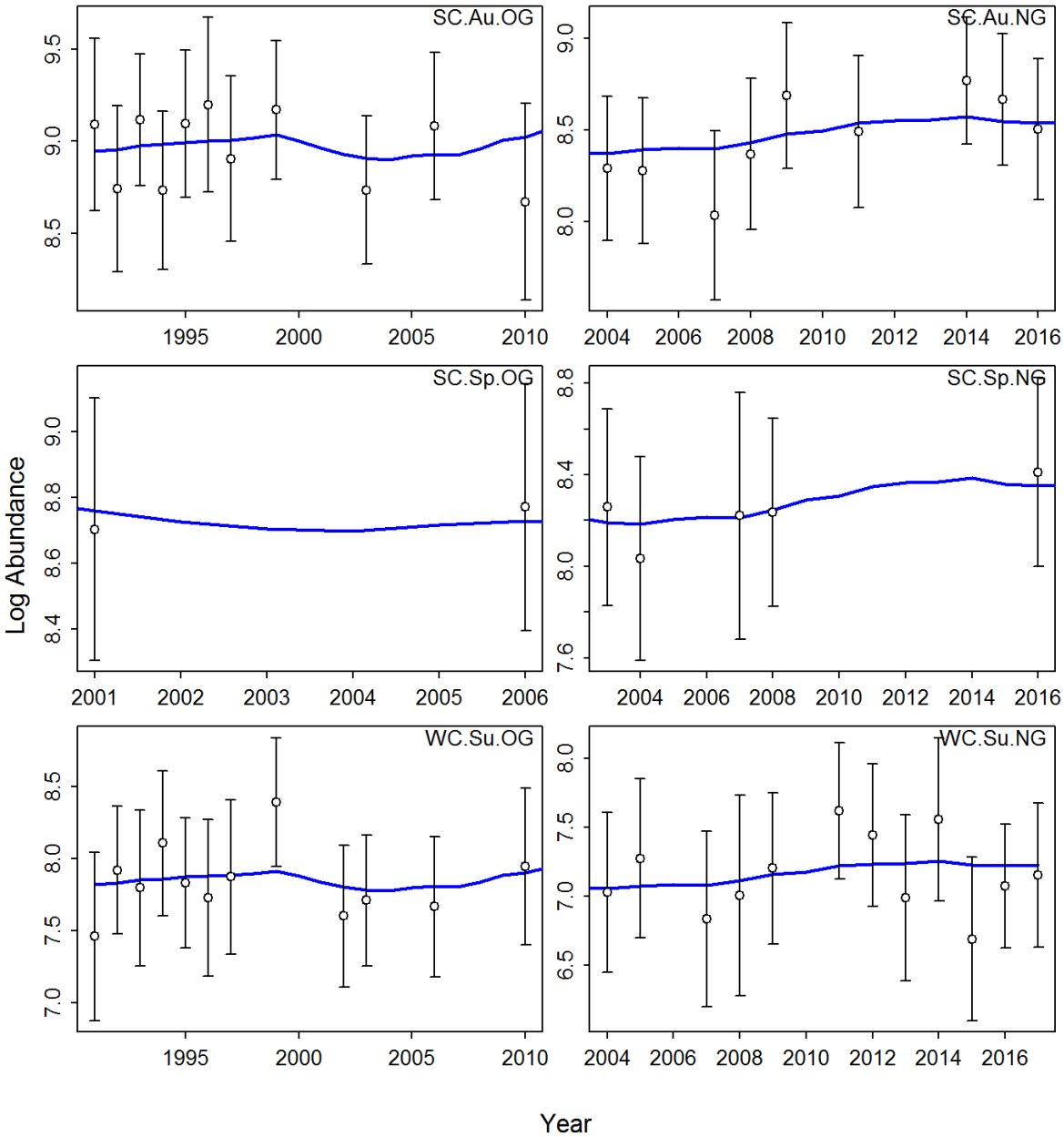

Figure 2: Model diagnostics plots produced by JARA using data from Whitespot Smoothhound *Mustelus palumbes*. Each panel shows the state-space model fit (blue line) to the natural logarithm (log) of the observed relative abundance data (white points), shown with 95% credible intervals (black lines) derived from the observation variance.

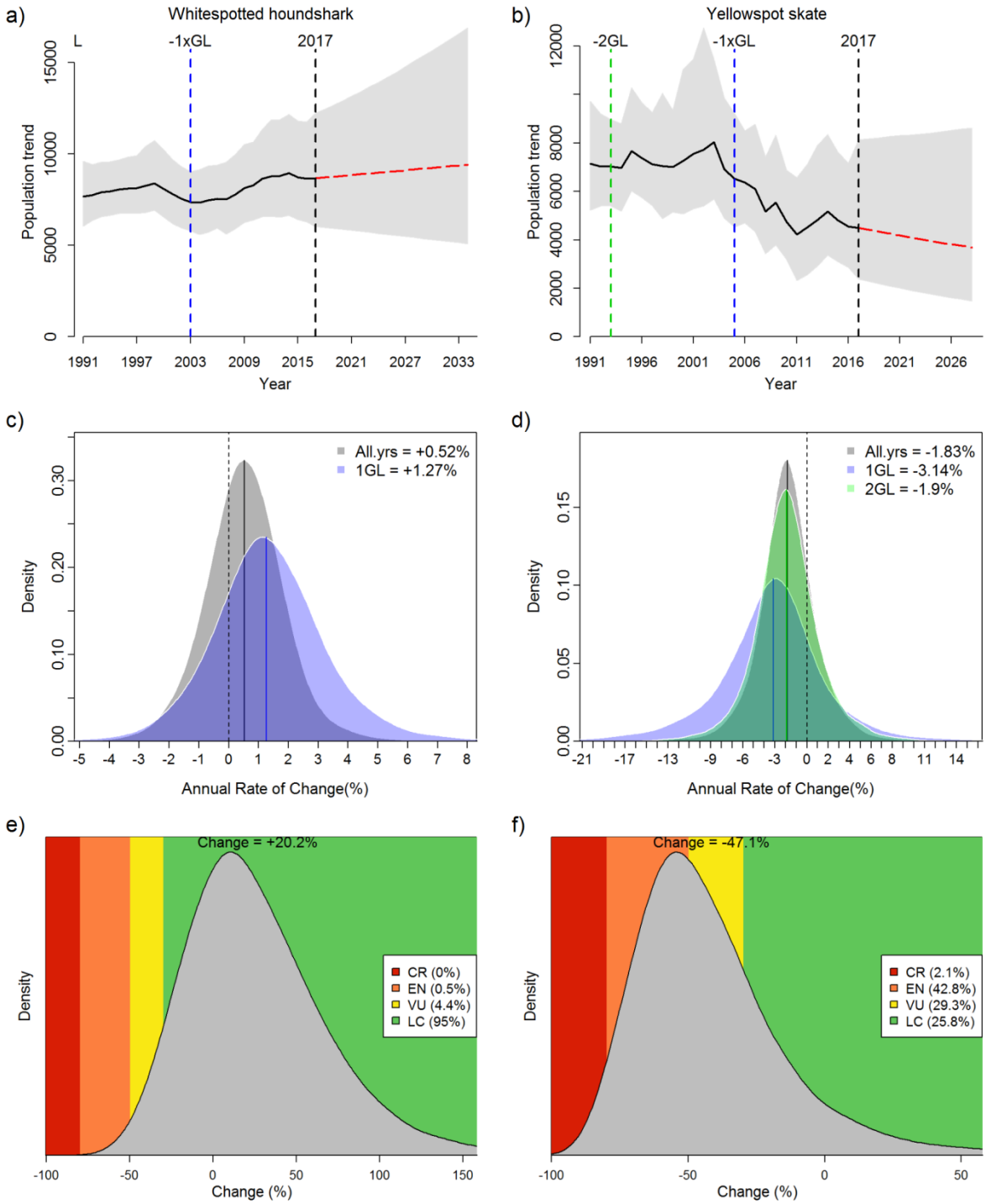

Figure 3: JARA decision-support plots for (left) Whitespot Smoothhound *Mustelus palumbes* and (right) Yellowspotted Skate *Leucoraja wallacei* showing (a) – (b) the overall JARA fit (black line) to the observed timeseries and the projected (red line) population trajectory over three generation length (GL); (c) – (d) the posterior probability for the percentage annual population change calculated from all the observed data (in black), from the last 1 generation length (in blue), from the last 2 generation lengths (in green) with the medians (solid lines) shown relative to a stable population (% change = 0, black dashed line); (e) – (f) the median change over three generation lengths ("Change = xx") and corresponding probabilities for rates of population change falling within the IUCN Red List categories.

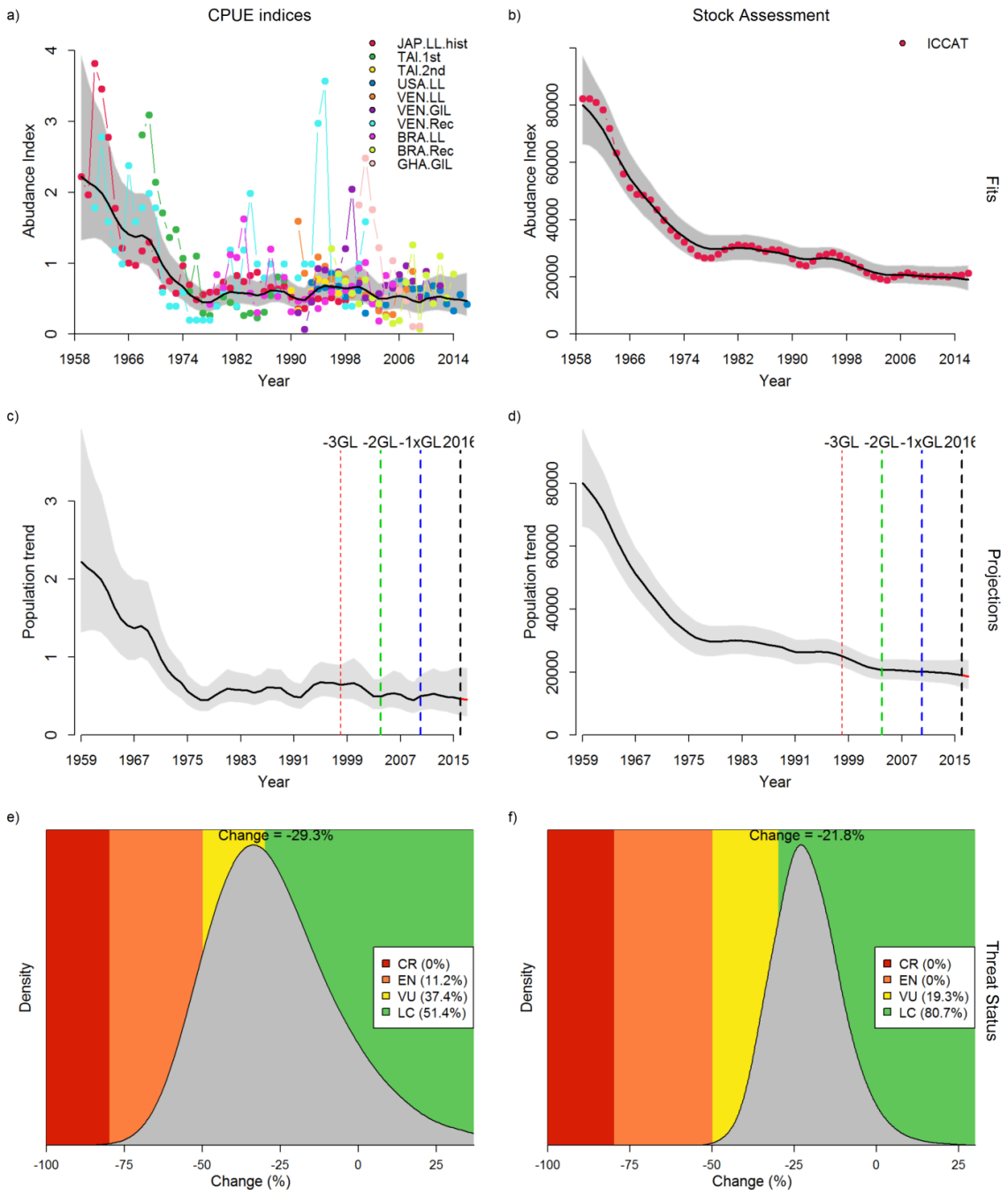

Figure 4: JARA plots for trend analyses of two types of abundance information for Blue marlin *Makaira nigricans*base: (left) ten standardized catch per unit effort (CPUE) indices from multiple fishing fleets (1959–2016) and (right) estimated biomass estimates from the 2018 Atlantic stock assessment conducted by the International Commission for the Conservation of Atlantic Tuna (ICCAT), showing (a) – (b) scaled observed relative abundance with model fits (predicted relative abundance, black line) and 95% posterior predictive credibility intervals (grey polygons); (c) – (d) the overall JARA fit (black line) over the observed timeseries with vertical dashed lines denoting the periods corresponding to 1–3 GL; (e) – (f) the median change over three generation lengths ("Change = xx") and corresponding probabilities for rates of population change falling within the IUCN Red List categories.

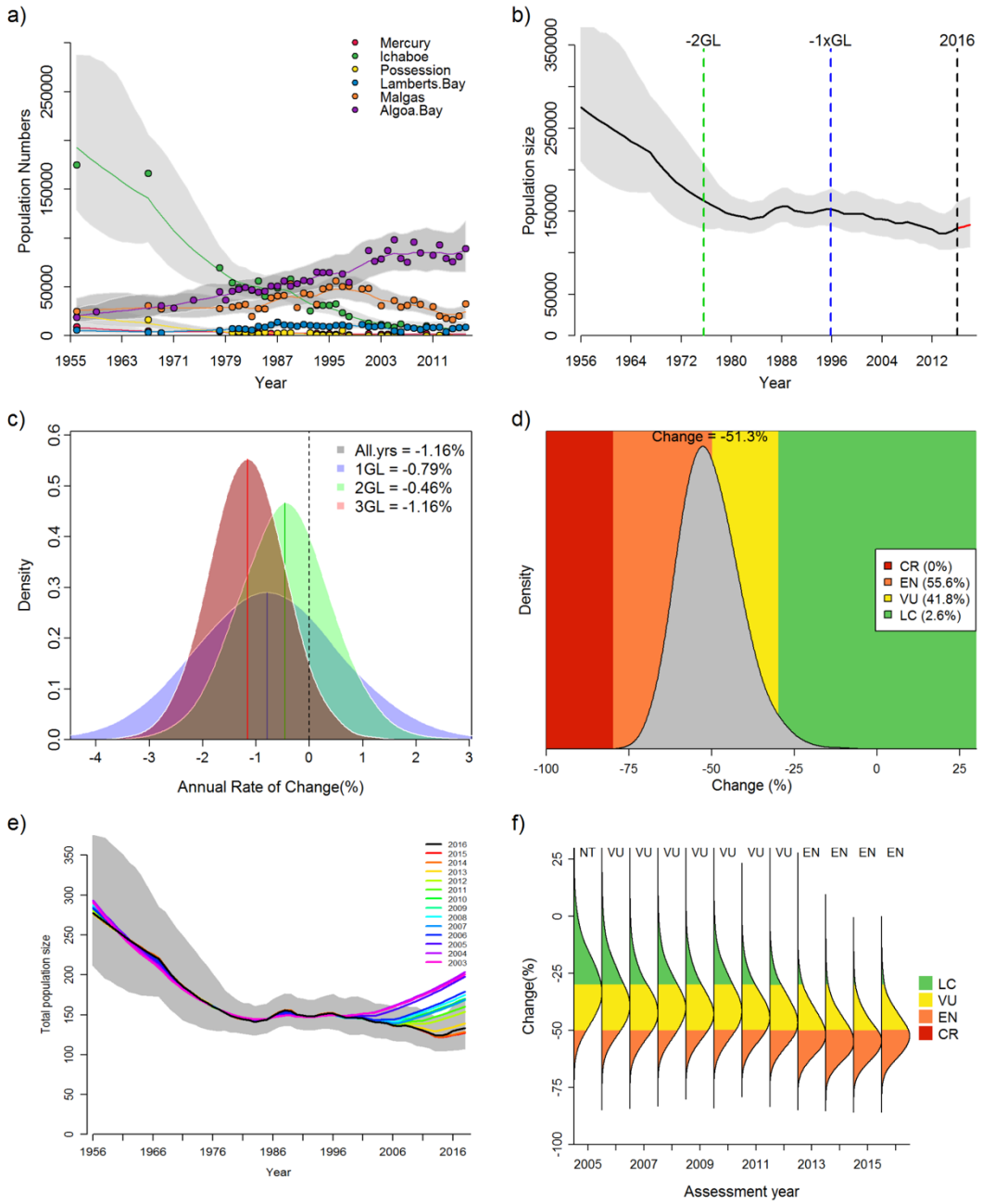

Figure 5: JARA decision-support plots for Cape Gannet *Morus capensis* showing (a) the JARA fits (coloured lines) and 95% credible intervals (grey polygons) to each observed abundance timeseries (coloured points); (b) the overall JARA fit (black line) to the observed timeseries and the projected (red line) population trajectory over three generation times (GLs); (c) the posterior probability for the percentage annual population change calculated from all the observed data (in black), from the last 1 generation length (in blue), from the last 2 generation lengths (in green), and from the last 3 generation lengths (in red) with the means (solid lines) shown relative to a stable population (% change = 0, black dashed line); (e) the median change over three GLs ("Change = xx") and corresponding probabilities for rates of population change falling within the IUCN Red List categories; (e) retrospective pattern of total abundance estimates obtained through sequential removal of terminal years and subsequent forward projections to attain 3 GL ("retrospective peel"); and corresponding retrospective status posteriors of change over three generation lengths - coloured according to the IUCN Red List category thresholds for the Red List criteria A2.

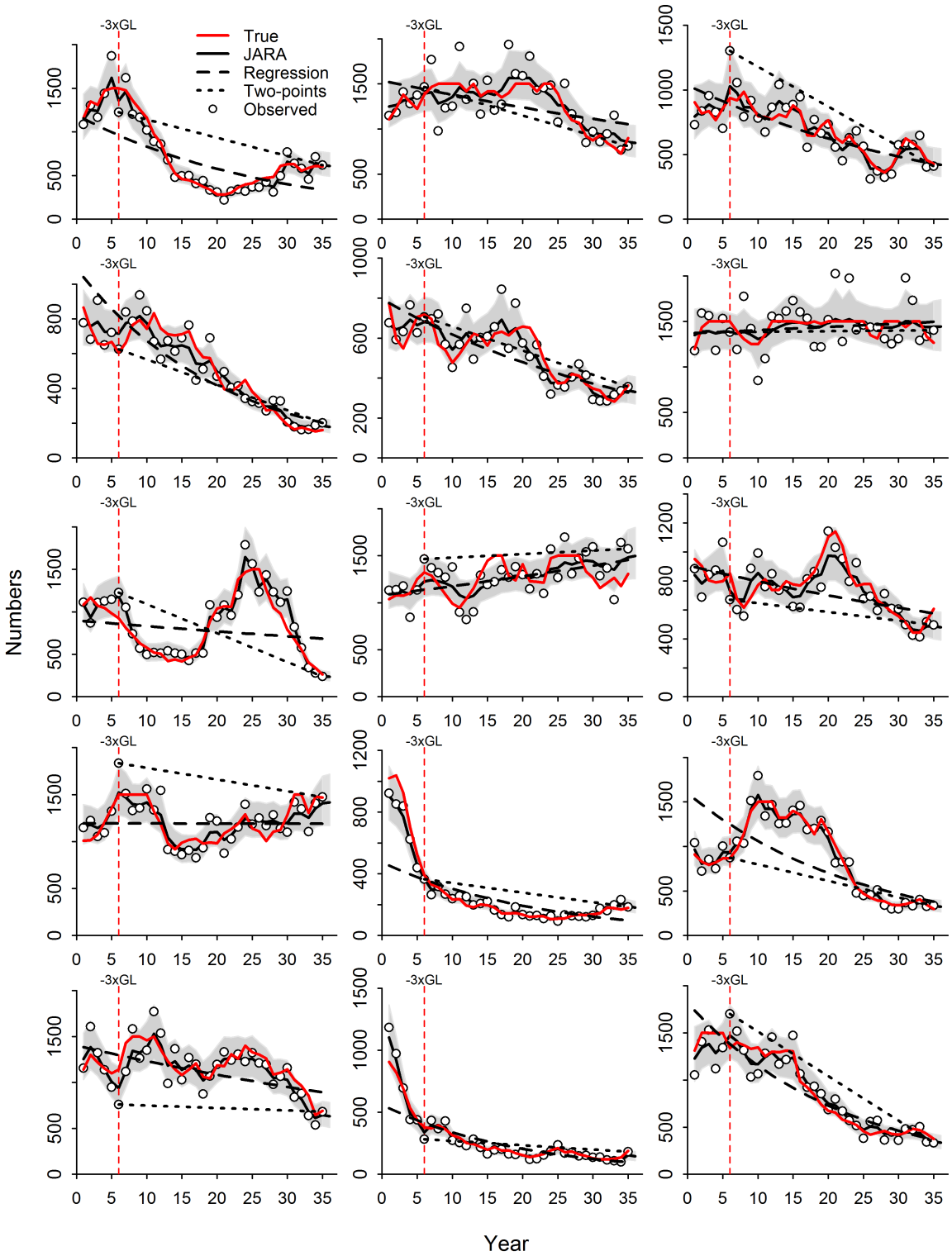

Figure. 6: Simulated 'true' abundance trajectories and observations for the first 15 of 200 simulation runs together with estimated abundance trends using three alternative estimation models (EMs): (1) 'Two-Point", (2) "Regression", and (3) 'JARA'. The vertical red dashed line denotes the assessment horizon of 30 years (3 GL) and grey polygons show the 95% credible intervals for the JARA fit.

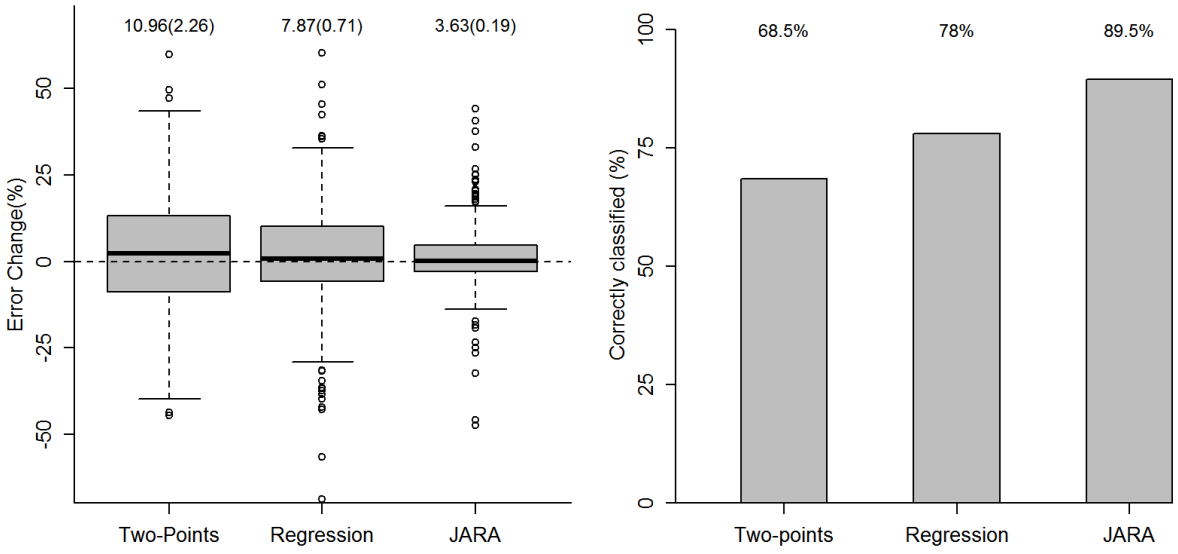

Figure. 7: Simulation testing results across 200 runs showing (a) boxplots of errors between the estimated and 'true' percentage change over 3 generation lengths (GL) with Median Absolute Errors and Median Errors (in brackets) displayed of each box and (b) barplots illustrating the percentages of correctly classified IUCN threat categories (LC, VU, EN, CR) for the three alternative estimation models (EMs): (1) 'Two-Point", (2) "Regression", and (3) 'JARA'.

References

Bauer, H., Chapron, G., Nowell, K., Henschel, P., Funston, P., Hunter, L.T.B., Macdonald, D.W. & Packer, C. (2015). Lion (*Panthera leo*) populations are declining rapidly across Africa, except in intensively managed areas. *Proceedings of the National Academy of Sciences*, 112, 14894–14899.

BirdLife-International. (2018). Morus capensis. The IUCN Red List of Threatened Species: e.T2269666. IUCN Red List. URL http://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22696668A132587992.en. Downloaded on 12 August 2019

Boyd, C., DeMaster, D.P., Waples, R.S., Ward, E.J. & Taylor, B.L. (2017). Consistent Extinction Risk Assessment under the U.S. Endangered Species Act. *Conservation Letters*, 10, 328–336.

Brodziak, J. & Ishimura, G. (2012). Development of Bayesian production models for assessing the North Pacific swordfish population. *Fisheries Science*, 77, 23–34.

Buckland, S.T., Newman, K.B., Thomas, L. & Koesters, N.B. (2004). State-space models for the dynamics of wild animal populations. *Ecological Modelling*, 171, 157–175.

Butchart, S.H.M., Stattersfield, A.J., Bennun, L.A., Shutes, S.M., Akçakaya, H.R., Baillie, J.E.M., Stuart, S.N., Hilton-Taylor, C. & Mace, G.M. (2004). Measuring global trends in the status of biodiversity: Red list indices for birds. *PLoS Biology*, 2, e383.

Carvalho, F., Ahrens, R., Murie, D., Ponciano, J.M., Aires-da-silva, A., Maunder, M.N. & Hazin, F. (2014). Incorporating specific change points in catchability in fisheries stock assessment models: An alternative approach applied to the blue shark (*Prionace glauca*) stock in the south Atlantic Ocean. *Fisheries Research*, 154, 135–146.

Caughley, G. (1994). Directions in Conservation Biology. *Journal of Animal Ecology*, 63, 215–244.

Chaloupka, M. & Balazs, G. (2007). Using Bayesian state-space modelling to assess the recovery and harvest potential of the Hawaiian green sea turtle stock. *Ecological Modelling*, 205, 93–109.

Connors, B.M., Cooper, A.B., Peterman, R.M. & Dulvy, N.K. (2014). The false classification of extinction risk in noisy environments. *Proceedings of the Royal Society B: Biological Sciences*, 281, 20132935.

Crawford, R.J.M., Dundee, B.L., Dyer, B.M., Klages, N.T., Meyër, M.A. & Upfold, L. (2007). Trends in numbers of Cape gannets (*Morus capensis*), 1956/57–2005/06, with a consideration of the influence of food and other factors. *ICES Journal of Marine Science*, 64, 169–177.

Crawford, R.J.M., Sydeman, W.J., Thompson, S.A., Sherley, R.B. & Makhado, A.B. (2019). Food habits of an endangered seabird indicate recent poor forage fish availability off western South Africa. *ICES Journal of Marine Science*, fsz081.

D'Eon-Eggertson, F., Dulvy, N.K. & Peterman, R.M. (2015). Reliable Identification of Declining Populations in an Uncertain World. *Conservation Letters*, 8, 86–96.

Distiller, G., Altwegg, R., Crawford, R.J.M., Klages, N.T.W. & Barham, B. (2012). Factors affecting adult survival and inter-colony movement at the three South African colonies of Cape gannet. *Marine Ecology Progress Series*, 461, 245–255.

Dulvy, N.K., Jennings, S., Rogers, S.I. & Maxwell, D.L. (2006). Threat and decline in fishes: an indicator of marine biodiversity. *Canadian Journal of Fisheries and Aquatic Sciences*, 63, 1267–1275.

Fox, R., Harrower, C.A., Bell, J.R., Shortall, C.R., Middlebrook, I. & Wilson, R.J. (2018). Insect population trends and the IUCN Red List process. *Journal of Insect Conservation*, 23, 269–278.

Francis, R.I.C.C., Hurst, R.J. & Renwick, J.A. (2003). Quantifying annual variation in catchability for commercial and research fishing. *Fishery Bulletin*, 101, 293–304.

Froese, R., Demirel, N., Coro, G., Kleisner, K.M. & Winker, H. (2017). Estimating fisheries reference points from catch and resilience. *Fish and Fisheries*, 18, 506–526.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*. (eds. Berger, J.O., Bernardo, J.M., Dawid, A.P. & Smith, A.F.M.). Clarendon Press, Oxford, pp. 169–193.

Heidelberger, P. & Welch, P.D. (1992). Simulation run length control in the presence of an initial transient. *Operations Research*, 31, 1109–1144.

Hilborn, R. & Mangel, M. (1997). *The Ecological Detective: Confronting Models with Data*. Princeton University Press, Princeton, NJ.

Hoffmann, M., Brooks, T.M., Da Fonseca, G.A.B., Gascon, C., Hawkins, A.F.A., James, R.E., Langhammer, P., Mittermeier, R.A., Pilgrim, J.D., Rodrigues, A.S.L. & Silva, J.M.C. (2008). Conservation planning and the IUCN Red List. *Endangered Species Research*, 6, 113–125.

IUCN. (2017). *Guidelines for using the IUCN Red List Categories and Criteria*. Version 13. Prepared by the Standards and Petitions Subcommittee. IUCN, Gland, Switzerland and Cambridge, UK.

Jiao, Y., Hayes, C. & Corte, E. (2009). Hierarchical Bayesian approach for population dynamics modelling of fish complexes without species-specific data. *ICES Journal of Marine Science*, 66, 367–377.

Keith, D., Akçakaya, H.R., Butchart, S.H.M., Collen, B., Dulvy, N.K., Holmes, E.E., Hutchings, J.A., Keinath, D., Schwartz, M.K., Shelton, A.O. & Waples, R.S. (2015). Temporal correlations in population trends: Conservation implications from time-series analysis of diverse animal taxa. *Biological Conservation*, 192, 247–257.

Kéry, M. & Schaub, M. (2012). *Bayesian population analysis using WinBUGS: A hierarchical perspective*. Academic Press, Oxford.

Lee, C.K.F., Keith, D.A., Nicholson, E. & Murray, N.J. (2019). REDLISTR: Tools for the IUCN Red Lists of Ecosystems and Threatened Species in R. *Ecography*, 42, 1–6.

Mace, G.M., Collar, N.J., Gaston, K.J., Hilton-Taylor, C., Akçakaya, H.R., Leader-Williams, N., Milner-Gulland, E.J. & Stuart, S.N. (2008). Quantification of extinction risk: IUCN's system for classifying threatened species. *Conservation Biology*, 22, 1424–1442.

Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models using Gibbs Sampling. 3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria.

Plummer, M., Nicky Best, Cowles, K. & Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6, 7–11.

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Regan, T.J., Taylor, B.L., Thompson, G.G., Cochrane, J.F., Ralls, K., Runge, M.C. & Merrick, R. (2013). Testing decision rules for categorizing species' extinction risk to help develop quantitative listing criteria for the U.S. endangered species act. *Conservation Biology*, 27, 821–831.

Rueda-Cediel, P., Anderson, K.E., Regan, T.J. & Regan, H.M. (2018). Effects of uncertainty and variability on population declines and IUCN Red List classifications. *Conservation Biology*, 32, 916–925.

Seminoff, J.A. & Shanker, K. (2008). Marine turtles and IUCN Red Listing: A review of the process, the pitfalls, and novel assessment approaches. *Journal of Experimental Marine Biology and Ecology*, 356, 52–68.

Sherley, R.B., Barham, B.J., Barham, P.J., Campbell, K.J., Crawford, R.J.M., Grigg, J., Horswill, C., Mcinnes, A., Morris, T.L., Pichegru, L., Steinfurth, A., Weller, F., Winker, H. & Votier, S.C. (2018). Bayesian inference reveals positive but subtle effects of experimental fishery closures on marine predator demographics. *Proceedings of the Royal Society of London, Series B: Biological Sciences,* 285, 20172443.

Simmons, R.E., Kolberg, H., Braby, R. & Erni, B. (2015). Declines in migrant shorebird populations from a winter-quarter perspective. *Conservation Biology*, 29, 877–887.

Su, Y.-S. & Yajima, M. (2012). *R2jags: A Package for Running jags from R*.

Thorson, J.T., Jensen, O.P. & Zipkin, E.F. (2014). How variable is recruitment for exploited marine fishes? A hierarchical model for testing life history theory. *Canadian Journal of Fisheries and Aquatic Sciences*, 71, 973–983.

Thorson, J.T. & Minto, C. (2015). Mixed effects: a unifying framework for statistical modelling in fisheries biology. *ICES Journal of Marine Science*, 72, 1245–1256.

de Valpine, P. & Hastings, A. (2002). Fitting population models incorporating process noise and observation error. *Ecological Monographs*, 72, 57–76.

Wilson, H.B., Kendall, B.E. & Possingham, H.P. (2011). Variability in Population Abundance and the Classification of Extinction Risk. *Conservation Biology*, 25, 747–757.

Winker, H., Carvalho, F. & Kapur, M. (2018). *JABBA: Just Another Bayesian Biomass Assessment. Fisheries Research*, 204, 275–288.

Supplementary Information

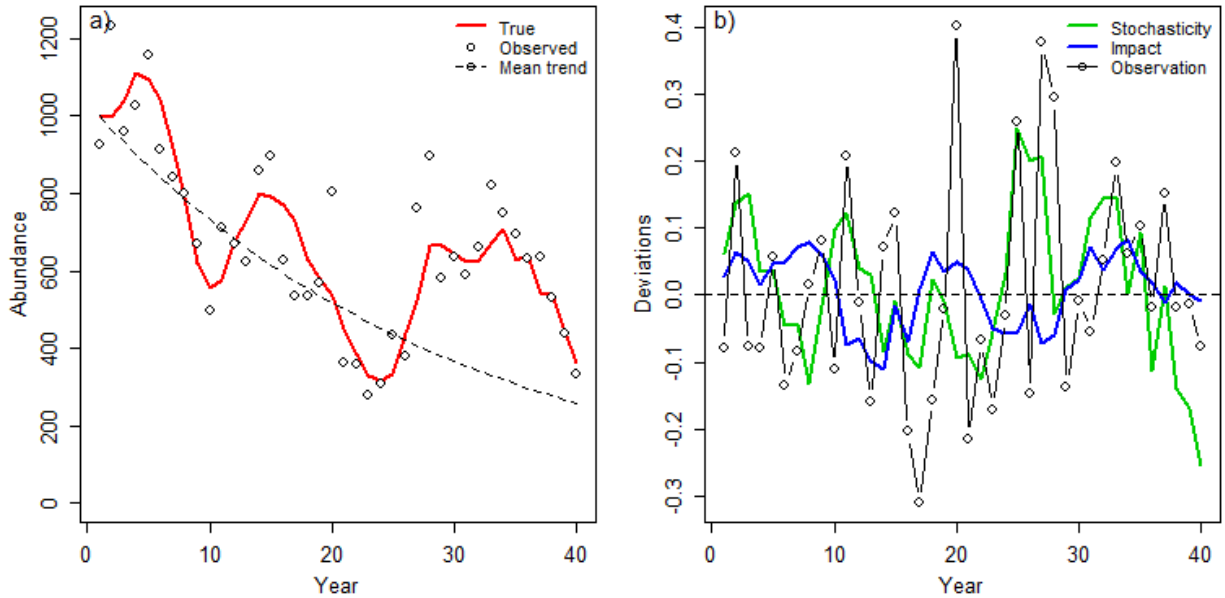

Figure S1: Illustrations of (a) a simulated 'true' population trajectory, the corresponding observed values given the observation error and the underlying deterministic mean trend $\bar{r}$; and (b) decomposed first-order autoregressive process error deviations simulating the natural variation in population numbers (Stochasticity) and a more persistent, regime-like variation in anthropogenic impact (Impact), together with a random observation deviations from the 'true' abundance due to random sampling error.

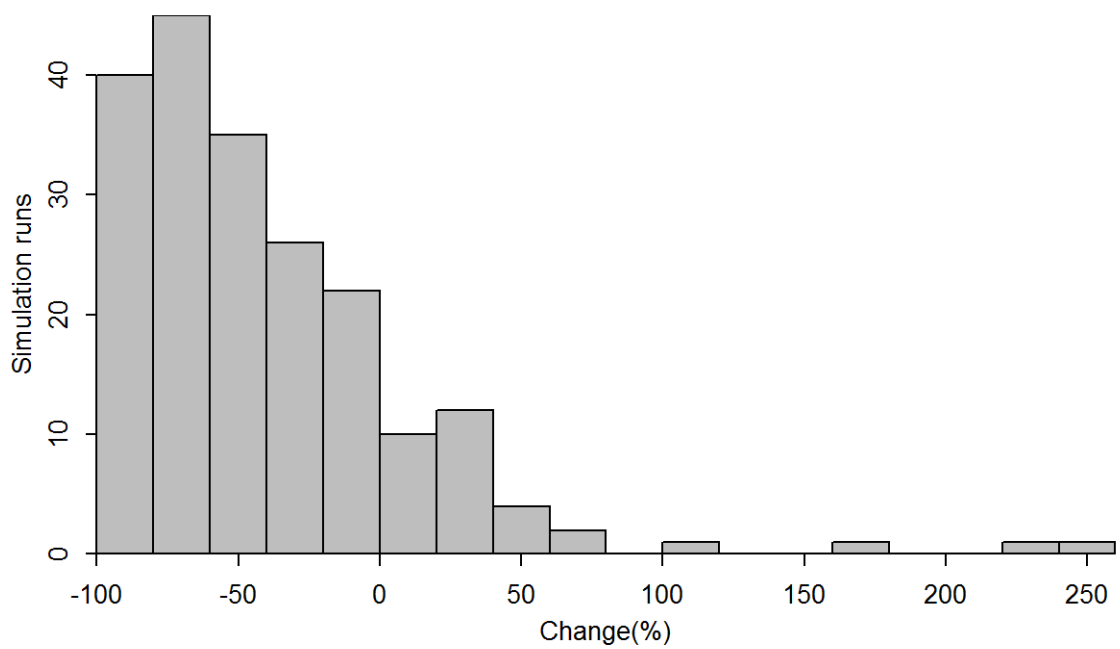

Figure S2: Simulated distribution of the 'true' %*C* generated from 200 simulation replicates of abundance trajectories over a 30-year assessment horizon.

```
#-----------------------------------------------------------------------------------------------------------
# Generic Population Simulator
#-----------------------------------------------------------------------------------------------------------

popsim = function(n0 = 1000 ,simyrs = 40,r.mu = runif(1,-0.05,0),proc.pop =
        0.1,AR1.pop=0.3,proc.imp=0.05,AR1.imp=0.8,CV.obs=0.15,GT=10,Plot=TRUE){
# Process error
# (1) stochastic variation in popdyn
pop_var = rnorm(simyrs,0,proc.pop)
# Add AR1 (serial auto-correlation)
pop_dev = pop_var[1]
for(y in 2:simyrs) pop_dev[y] =  AR1.pop * pop_dev[y-1]+pop_var[y]*sqrt(1-AR1.pop^2)

# (2) random variation in impact (e.g. offtake)
imp_var = rnorm(simyrs,0,proc.imp)
# Add AR1 (serial auto-correlation)
imp_dev = imp_var[1]
for(y in 2:simyrs) imp_dev[y] =  AR1.imp * imp_dev[y-1]+imp_var[y]*sqrt(1-AR1.imp^2)

# Combine all as latent effect in r vector given r.mu
r = r.mu+pop_dev-imp_dev

# Add random observation error
obs_dev = rnorm(simyrs,0,CV.obs)

# True Population
n_t = n0*exp(r[1])
for(t in 2:simyrs){
n_t[t] =  n_t[t-1]*exp(r[t-1])
if(n_t[t]>n0*1.5) n_t[t] = n0*1.5}

# Deterministic trend
nbar = n0
for(t in 2:simyrs){
nbar[t] =  nbar[t-1]*exp(r.mu)}

# Imperfect observation
y_t =exp(log(n_t)+obs_dev)

if(Plot==TRUE){
Par = list(mfrow=c(1,2),mar = c(4, 4, 1, 1), mgp =c(2.5,1,0),mai = c(0.6, 0.6, 0.1, 0.1),mex=0.8, tck = -0.02,cex=0.8)
par(Par)
# Error
plot(n_t,ylim=c(0,max(n_t,y_t,n0*1.05)),type="l",col=2,lwd=2,ylab="Abundance",xlab="Year")
points(y_t,pch=21,bg="white",type="p")
lines(nbar,lty=2)
legend("topright",c("True","Observed","Mean trend"),pch=c(-1,1,1),lwd=c(2,-1,1),col=c(2,1,1),lty=c(1,-1,2),bty="n",cex=0.8)
legend("topleft","a)",bty="n",cex=1.1,x.intersp = -0.5,y.intersp = -0.2)
plot(pop_dev,ylim=range(c(pop_dev,imp_dev,obs_dev)),type="l",col=3,lwd=2,ylab="Deviations",xlab="Year")
lines(imp_dev,col=4,lwd=2)
legend("topright",c("Stochasticity","Impact","Observation"),pch=c(-1,-1,1),lwd=c(2,2,1),col=c(3,4,1),bty="n",cex=0.8)
points(obs_dev,type="b")
abline(h=0,lty=2)
legend("topleft","b)",bty="n",cex=1.1,x.intersp = -0.5,y.intersp = -0.2)

}
return(list(dat=data.frame(yr=1:simyrs,y=y_t),N=data.frame(yr=1:simyrs,N=n_t),r.mu=r.mu,perc.change=(n_t[simyrs]/n_t[simyrs-3*GT+1]-
        1)*100))
}

# Run function
popsim()
```