JARA: 'Just Another Red-List Assessment'

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

[1] Joint Research Centre (JRC), European Commission, Via E. Fermi 2749, 21027 Ispra, Italy.

[2] Earth to Ocean Research Group, Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A1S6, Canada.

[3] Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Penryn Campus, Cornwall, TR10 9FE, UK.

*Correspondence: Henning Winker. E-mail: henning.winker@ec.europa.eu; 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 according to population reduction thresholds. 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 is implemented via an R package that 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. JARA outputs display easy to interpret graphics of the posterior probability of the population trend against the corresponding IUCN Red List categories, as well as additional graphics to describe the timeseries and model diagnostics. 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, with substantial benefits for global conservation efforts.

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 [1,2]. Although a number classification protocols exist to assess a species' extinction risk [3], the International Union for the Conservation of Nature's (IUCN) Red List of Threatened Species is viewed widely as the global standard [1,4].

To list a species in a threatened category (Critically Endangered [CR], Endangered [EN] or Vulnerable [VU]) on the Red List, expert assessors consider the risks associated with both the small-population paradigm and the declining population paradigm [5] 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 [4,6]. 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 [4]. 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 [7]. 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 [2,4,8–10].

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 [4,6]. Estimating the decline between just two points in time remains widely applied as a rapid assessment approach [6,11,12]. 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 span the required IUCN assessment time frame of the longest period between three generation lengths (GL) or 10 years 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 [11,12]. Fitting linear or log-linear regression models to timeseries of abundance observations is a second widely-applied approach [6,8,10,11]. 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 [6]. 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 [13]. 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 [14–16].

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 [17,18]. 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 [18]. 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 timeseries data. Given these desirable properties, it is not surprising that there has been a rapid uptake of state-space approaches for modelling population dynamics [17,19,20], including more recent applications of quantifying population decline [6,21–23]. 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 [3,6,16].

Bayesian implementations of state-space models offer a powerful framework to improve the characterization and communication of uncertainty during IUCN Red List assessments [24]. The posterior probabilities provide an intuitive and transparent way to express uncertainty about population declines to conservation practitioners [22,25], which can be translated directly into probabilistic statements about a population falling into a threatened category [22,23]. However, developing customized Bayesian state-space models can be technically demanding and time consuming, especially when dealing with often case-specific, 'noisy' 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 'Just Another Red-List

Assessment' (JARA). 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, reproducibility and to allow further development, we provide the R package JARA on the global open-source platform GitHub (https://github.com/henning-winker/JARA), so that JARA can be readily 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.

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', [26]), 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, [27]). JARA enables analysis of one or multiple abundance indices simultaneously, where each index can have a different scale, 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 of JARA

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 $I_{t+1} = I_t\lambda_t$, where λ_t is the growth rate in year *t* [18]. 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 [21]. 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). Because the process error is lognormally distributed, r_t is adjusted for lognormal bias by subtracting half the variance (otherwise the stochastic lognormal error induces a small positive bias) at each time step: $r_t = \bar{r}_p + r_{t-1} - (0.5 * \sigma_\eta^2)$ [28]. The corresponding observation equation is of the form $\log(y_t) = \mu_t + \epsilon_t$, where y_t denotes the abundance observation 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 [27] 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 timeseries of absolute abundance observations, 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: $r_{t,i} \sim Normal(\bar{r}_i, \sigma_\eta^2)$. Again, we adjusted $r_{t,i}$ for lognormal bias by subtracting half the process variance at each time step: $r_t = \bar{r}_i + r_{t-1,i} - (0.5 * \sigma_\eta^2)$ [28]. 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 terms $\varepsilon_{t,i} \sim Normal(0, \sigma_{\varepsilon_i}^2)$ now become the specific to subpopulation *i.* The estimated posterior of the global population trajectory $\hat{I}_{p,t}$ for year *t* is then computed from the sum of all individual subpopulation trajectory posteriors, after correcting them for the lognormal bias by subtracting half their variance [28,29]: $\hat{I}_{p,t} =$

$\sum_i \exp(\mu_{t,i} - 0.5 * \sigma^2_{\mu_{t,i}})$. Although the sum of multiple lognormal distributions are not themselves lognormally distributed [30], such that the posterior median or uncorrected mean of $\hat{I}_{p,t}$ would yield inflated estimates (particularly) when population counts are large [29], our corrected mean is unbiased [31].

Bayesian Framework

JARA is run from the statistical environment R [32] and executed in JAGS [26], using a wrapper function from the R library 'r2jags' [33]. The Bayesian posterior distributions are estimated by means of Markov Chain Monte Carlo (MCMC) simulation. In JAGS, all estimable hyper-parameters are 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 [34], 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. Vague normal priors with a mean of 0 and variance of 1000, $Normal(0,1000)$, 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 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 (σ^2_η) 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: $\sigma^2 \sim 1/\text{gamma}(0.001,0.001)$ [35–37], which yields an approximately uniform prior on the log scale [27].

The total observation variance σ^2_ε can be separated into three components: (1) externally derived standard errors $\sigma^2_{SE_{y,i}}$ for each abundance index *i*, (2) a fixed input variance σ^2_{fix} 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 [38], so the total observation variance for abundance index *i* and year *y* is: $\sigma_{\varepsilon_{y,i}}^2 = \sigma_{SE_{y,i}}^2 + \sigma_{fix}^2 + \sigma_{est}^2$ [39]. Each variance component can be individually switched on or off by the user [27]. Adding a fixed observation error is common practice to account for additional sampling error associated with abundance indices [40]. Setting a minimum plausible observation error in this way informs the estimate of the process variance as a portion of total variance is assigned *a priori* to the observation variance [27] and helps to increase model stability and convergence of state-space models [41].

<u>*Estimating probabilities of population decline*</u>

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 GL, *C%* is automatically calculated as the difference between the median of three years around the final observed data point T, and a three-year median 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 [42]. 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 timeseries: $\tilde{r}_i = \frac{1}{T}\sum_{t=1}^{T} r_{t,i}$.

<u>*Model Diagnostics*</u>

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 timeseries 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 timeseries [27]. 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 [43], adopting minimal thresholds of p = 0.05 for Heidelberger and Welch [44] and Geweke's [45] diagnostics called from R. Unless otherwise specified, JARA reports medians and 95% credible intervals (CI).

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 timeseries 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 timeseries), 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 (Critically Endangered CR, Endangered EN, Vulnerable VU , Near Threatened NT or Least Concern LC) 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 [11]. 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 [15,16] 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 [46]. 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 timeseries 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 and Discussion

<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

303 Whitespot Smoothhound *Mustelus palumbes*; (2) a comparison of model-predicted biomass
304 trajectories from the 2018 Atlantic Blue Marlin *Makaira nigricans* stock assessments and
305 standardized catch-per-unit-effort (CPUE) abundance indices from various fishing fleet, which
306 were used as input into the assessment model and (3) census data for a large seabird, the Cape
307 Gannet *Morus capensis*, in the form of total breeding size estimates for its six global breeding
308 localities in Namibia and South Africa (1979–2017). All worked examples and input data are
309 available on GitHub (https://github.com/henning-winker/JARA).
310
311 (1) Whitespot Smoothhound and Yellowspotted Skate are both endemic to southern Africa and
312 had been previously listed as Data Deficient by the IUCN [47,48]. During the 2018 Southern
313 African IUCN Red List workshop, hosted by the IUCN Shark Specialist Group (SSG), both species
314 were reassessed using JARA as a decision-support tool. Based on available life-history
315 information, the GLs were empirically estimated as 14 years for Whitespot smoothhound and 12
316 years for Yellowspotted Skate. Long-term abundance indices were obtained from analysis of
317 demersal research trawl survey data that were conducted with alternating trawl gear configurations
318 (old/new gear) during summer along South Africa's west coast (1991–2017) and during autumn
319 and spring along the south coast (1991–2016) by the Department of Agriculture, Forestry and
320 Fisheries. Yellowspotted Skate was only encountered during South Coast Surveys. Annual density
321 estimates (kg per nm^2 area swept) and associated standard errors on log-scale were estimated using
322 a geostatistical delta-GLMM [49,50]. We only considered the period from 1991 onwards due to
323 evident learning effects for identifying Chondrichthyes during initial survey years. The estimated
324 abundance indices were separately derived for each gear type, survey (coast/season). The input
325 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 [51]. 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). 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 [52]. 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 [53]. Counts were sporadic until 1978/79 but undertaken approximately annually thereafter (Figure 5). Cape Gannets generally have high adult survival, strong breeding site fidelity [54] and low reproductive success (≤ 0.82 chicks fledged per pair annually; [55]. Accordingly, their GL is estimated as 20.2 years [56] 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

395 was ~3.7 times larger for the regression EM (0.71%) and ~12 times larger for the Two-Point EM
396 (2.26%) than for JARA (0.19%), which was approximately unbiased. Moreover, the median %*C*
397 from JARA correctly identified the 'true' IUCN threat category in 89.5% of cases; > 10% more
398 than the regression EM (which was based on the current IUCN guidelines) and >20% more than
399 the Two-Point EM (Figure 7). Although these might seem like relatively small improvements,
400 misclassification on the IUCN Red List can have real consequences for species conservation
401 [7,57], particularly resource allocation if species move in and out of the threatened categories [58].
402 There will be many real-world cases where the observed or estimated decline is close to a Red List
403 category threshold (e.g. >50% decline for EN), as was the case for Yellowspotted Skate (median
404 %*C* −47.1%, Figure 3f) and Cape Gannet (median %*C* −51.3%, Figure 5d) in our examples. Not
405 only would we would expect JARA to indicate the 'true' IUCN threat category more reliably in
406 these sorts of situations, but the additional information provided by JARA – particularly the plot
407 of the probabilities that the rate of change falls within the IUCN Red List categories (e.g. Figure
408 5d) – offers assessors insight into the weight of evidence underpinning their assignment.
409

410 *Current limitations and real-world applications*

411 Although still being refined, JARA has already been used to support expert decision making in the
412 assessment of a terrestrial mammal [59], a number of elasmobranchs [31,60] and two seabirds
413 [31,53]. We designed JARA with the aim of supporting the global effort to provide robust,
414 consistent, and transparent assessments for all species under the IUCN Red List criteria [24]. And
415 we designed it to be free, open-source, and easy to use, to support access for conservation
416 practitioners, researchers and assessors working in places with limited opportunities to access
417 proprietary software [12]. While we certainly hope that more conservation practitioners will use

JARA, we caution against applying it to unfamiliar timeseries or blindly accepting the status suggested by decision-support tools [12]. Undertaking Red List assessments requires assessors to understand their target species and the rules specified in the Red List guidelines. The data used for assessments should be proofed and verified by specialists (e.g. an IUCN Specialist Group) to ensure their reliability, users should make informed ecological judgments regarding the length of the timeseries relative to a species' GL when assessing against criterion A, and should remain cognisant of biases in data availability or survey effort when making their recommendations [11,12,24]. Moreover, JARA has been designed to be used as a decision-support tool and its outputs are not necessarily indicative of a final classification of extinction risk [24]. Rather, assessors should use tools like JARA to make an informed interpretation based on all of the Red List guidelines before coming to a decision on the final assessment outcome [12,24]. Nevertheless, when coupled with expert oversight, JARA has wide potential applicability to other studies of wildlife populations threatened with extinction [61,62].

Acknowledgments

H.W. and R.B.S. thank Marc Kéry and Michael Schaub for introducing us to Bayesian state-space models, and we thank Nick Dulvy for helpful conversations about the Red List process during the development of JARA. RBS was supported by the Pew Fellows Program in Marine Conservation at The Pew Charitable Trusts. The views expressed are those of the authors and do not necessarily reflect the views of The Pew Charitable Trusts. N.P. was funded by the Shark Conservation Fund, a philanthropic collaborative pooling expertise and resources to meet the threats facing the world's sharks and rays. The Shark Conservation Fund is a project of Rockefeller Philanthropy Advisors.

441 **References**

442 1. Hoffmann M, Brooks TM, Da Fonseca GAB, Gascon C, Hawkins AFA, James RE, et al.
443 Conservation planning and the IUCN Red List. Endanger Species Res. 2008; 6: 113–125.
444 doi:10.3354/esr00087

445 2. Butchart SHM, Stattersfield AJ, Bennun LA, Shutes SM, Akçakaya HR, Baillie JEM, et
446 al. Measuring Global Trends in the Status of Biodiversity: Red List Indices for Birds. PLoS
447 Biol. 2004; 2: e383. doi:10.1371/journal.pbio.0020383

448 3. Regan TJ, Taylor BL, Thompson GG, Cochrane JF, Ralls K, Runge MC, et al. Testing
449 decision rules for categorizing species' extinction risk to help develop quantitative listing
450 criteria for the U.S. endangered species act. Conserv Biol. 2013; 27: 821–831.
451 doi:10.1111/cobi.12055

452 4. Mace GM, Collar NJ, Gaston KJ, Hilton-Taylor C, Akçakaya HR, Leader-Williams N, et
453 al. Quantification of extinction risk: IUCN's system for classifying threatened species.
454 Conserv Biology. 2008; 22: 1424–1442. doi:10.1111/j.1523-1739.2008.01044.x

455 5. Caughley G. Directions in Conservation Biology. J Anim Ecol. 1994; 63: 215–244.

456 6. Wilson HB, Kendall BE, Possingham HP. Variability in Population Abundance and the
457 Classification of Extinction Risk. Conserv Biol. 2011; 25: 747–757. doi:10.1111/j.1523-
458 1739.2011.01671.x

459 7. Rueda-Cediel P, Anderson KE, Regan TJ, Regan HM. Effects of uncertainty and variability
460 on population declines and IUCN Red List classifications. Conserv Biol. 2018; 32: 916–
461 925. doi:10.1111/cobi.13081

462 8. Dulvy NK, Jennings S, Rogers SI, Maxwell DL. Threat and decline in fishes: an indicator
463 of marine biodiversity. Can J Fish Aquat Sci. 2006; 63: 1267–1275. doi:10.1139/f06-035

464 9. Seminoff JA, Shanker K. Marine turtles and IUCN Red Listing: A review of the process,
465 the pitfalls, and novel assessment approaches. J Exp Mar Biol Ecol. 2008; 356: 52–68.
466 doi:10.1016/j.jembe.2007.12.007
467 10. Fox R, Harrower CA, Bell JR, Shortall CR, Middlebrook I, Wilson RJ. Insect population
468 trends and the IUCN Red List process. J Insect Conserv. 2018; 23: 269–278.
469 doi:10.1007/s10841-018-0117-1
470 11. IUCN Standards and Petitions Subcommittee. Guidelines for Using the IUCN Red List
471 Categories and Criteria. IUCN, Gland, Switzerland and Cambridge, UK; 2019.
472 12. Lee CKF, Keith DA, Nicholson E, Murray NJ. REDLISTR: Tools for the IUCN Red Lists
473 of Ecosystems and Threatened Species in R. Ecography. 2019; 42: 1–6.
474 doi:10.1111/ecog.04143
475 13. Hilborn R, Mangel M. The Ecological Detective: Confronting Models with Data.
476 Princeton, NJ: Princeton University Press; 1997.
477 14. D'Eon-Eggertson F, Dulvy NK, Peterman RM. Reliable Identification of Declining
478 Populations in an Uncertain World. Conserv Lett. 2015; 8: 86–96. doi:10.1111/conl.12123
479 15. Keith D, Akçakaya HR, Butchart SHM, Collen B, Dulvy NK, Holmes EE, et al. Temporal
480 correlations in population trends: Conservation implications from time-series analysis of
481 diverse animal taxa. Biol Conserv. 2015; 192: 247–257. doi:10.1016/j.biocon.2015.09.021
482 16. Connors BM, Cooper AB, Peterman RM, Dulvy NK. The false classification of extinction
483 risk in noisy environments. Proc Biol Sci. 2014; 281: 20132935.
484 doi:10.1098/rspb.2013.2935
485 17. de Valpine P, Hastings A. Fitting Population Models Incorporating Process Noise and
486 Observation Error. Ecol Monogr. 2002; 72: 57–76. doi:10.1890/0012-

9615(2002)072[0057:fpmipn]2.0.co;2

18. Kéry M, Schaub M. Bayesian population analysis using WinBUGS: A hierarchical perspective. Oxford: Academic Press; 2012.

19. Buckland ST, Newman KB, Thomas L, Koesters NB. State-space models for the dynamics of wild animal populations. Ecol Modell. 2004; 171: 157–175. doi:10.1016/j.ecolmodel.2003.08.002

20. Thorson JT, Minto C. Mixed effects: a unifying framework for statistical modelling in fisheries biology. ICES J Mar Sci. 2015; 72: 1245–1256. doi:10.1093/icesjms/fsu213. doi:10.1093/icesjms/fst048

21. Simmons RE, Kolberg H, Braby R, Erni B. Declines in migrant shorebird populations from a winter-quarter perspective. Conserv Biol. 2015; 29: 877–887. doi:10.1111/cobi.12493

22. Bauer H, Chapron G, Nowell K, Henschel P, Funston P, Hunter LTB, et al. Lion (Panthera leo) populations are declining rapidly across Africa, except in intensively managed areas. Proc Natl Acad Sci U S A. 2015; 112: 14894–14899. doi:10.1073/pnas.1500664112

23. Boyd C, DeMaster DP, Waples RS, Ward EJ, Taylor BL. Consistent Extinction Risk Assessment under the U.S. Endangered Species Act. Conserv Lett. 2017; 10: 328–336. doi:10.1111/conl.12269

24. Sherley RB, Winker H, Rigby CL, Kyne PM, Pollom R, Pacoureau N, et al. Estimating IUCN Red List population reduction: JARA—A decision-support tool applied to pelagic sharks. Conserv Lett. 2020; 13: e12688. doi:10.1111/conl.12688

25. Sherley RB, Barham BJ, Barham PJ, Campbell KJ, Crawford RJM, Grigg J, et al. Bayesian inference reveals positive but subtle effects of experimental fishery closures on marine predator demographics. Proc Biol Sci. 2018; 285: 20172443. doi:10.1098/rspb.2017.2443

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

27. Winker H, Carvalho F, Kapur M. JABBA: Just Another Bayesian Biomass Assessment. Fish Res. 2018; 204: 275–288. doi:10.1016/j.fishres.2018.03.010

28. Methot RD, Taylor IG. Adjusting for bias due to variability of estimated recruitments in fisheries assessment models. Can J Fish Aquat Sci. 2011; 68: 1744–1760.

29. Che-Castaldo C, Jenouvrier S, Youngflesh C, Shoemaker KT, Humphries G, McDowall P, et al. Pan-Antarctic analysis aggregating spatial estimates of Adélie penguin abundance reveals robust dynamics despite stochastic noise. Nat Commun. 2017; 8: 832. doi:10.1038/s41467-017-00890-0

30. Dufresne D. The log-normal approximation in financial and other computations. Adv Appl Probab. 2004; 36: 747–773.

31. Sherley RB, Crawford RJM, de Blocq AD, Dyer BM, Geldenhuys D, Hagen C, et al. The conservation status and population decline of the African penguin deconstructed in space and time. Ecol Evol. 2020; Forthcoming.

32. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2018. Available: http://www.r-project.org/

33. Su Y-S, Yajima M. R2jags: A Package for Running jags from R; 2012.

34. Jiao Y, Hayes C, Corte E. Hierarchical Bayesian approach for population dynamics modelling of fish complexes without species-specific data. ICES J Mar Sci. 2009; 66: 367–377. doi:10.1093/icesjms/fsn162

35. Chaloupka M, Balazs G. Using Bayesian state-space modelling to assess the recovery and

harvest potential of the Hawaiian green sea turtle stock. Ecol Modell. 2007; 205: 93–109. doi:10.1016/j.ecolmodel.2007.02.010

36. Brodziak J, Ishimura G. Development of Bayesian production models for assessing the North Pacific swordfish population. Fish Sci. 2012; 77: 23–34. doi:10.1007/s12562-010-0300-0

37. Carvalho F, Ahrens R, Murie D, Ponciano JM, Aires-da-Silva A, Maunder MN, et al. 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. Fish Res. 2014; 154: 135–146. doi:10.1016/j.fishres.2014.01.022

38. Francis RICC, Hurst RJ, Renwick JA. Quantifying annual variation in catchability for commercial and research fishing. Fish Bull. 2003; 101: 293–304.

39. Conn PB. Hierarchical analysis of multiple noisy abundance indices. Can J Fish Aquat Sci. 2010; 67: 108–120. doi:10.1139/F09-175

40. Maunder MN, Piner KR. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Fish Res. 2017; 192: 16–27.

41. Auger-Méthé M, Field C, Albertsen CM, Derocher AE, Lewis MA, Jonsen ID, et al. State-space models' dirty little secrets: even simple linear Gaussian models can have estimation problems. Sci Rep. 2016; 6: 26677. doi:10.1038/srep26677

42. Froese R, Demirel N, Coro G, Kleisner KM, Winker H. Estimating fisheries reference points from catch and resilience. Fish Fish. 2017; 18: 506–526. doi:10.1111/faf.12190

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

44. Heidelberger P, Welch PD. Simulation run length control in the presence of an initial transient. Oper Res. 1992; 31: 1109–1144. doi:10.1287/opre.31.6.1109

45. Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Berger JO, Bernardo JM, Dawid AP, Smith AFM, editors. Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting. Oxford: Clarendon Press; 1992. pp. 169–193.

46. Thorson JT, Jensen OP, Zipkin EF. How variable is recruitment for exploited marine fishes? A hierarchical model for testing life history theory. Can J Fish Aquat Sci. 2014; 71: 973–983.

47. Smale M. Mustelus palumbes. In: The IUCN Red List of Threatened Species 2006: e.T60247A12333813 [Internet]. 2006 [cited 13 Jun 2020]. Available: https://dx.doi.org/10.2305/IUCN.UK.2006.RLTS.T60247A12333813.en.

48. Smale M. Leucoraja wallacei. In: The IUCN Red List of Threatened Species 2009: e.T161492A5436168 [Internet]. 2009 [cited 13 Jun 2020]. Available: https://dx.doi.org/10.2305/IUCN.UK.2009-2.RLTS.T161492A5436168.en

49. Thorson JT, Shelton AO, Ward EJ, Skaug HJ. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES J Mar Sci. 2015; 72: 1297–1310. doi:10.1093/icesjms/fsu243

50. Currie JC, Thorson JT, Sink KJ, Atkinson LJ, Fairweather TP, Winker H. A novel approach to assess distribution trends from fisheries survey data. Fish Res. 2019; 214: 98–109. doi:10.1016/j.fishres.2019.02.004

51. Collette B, Acero A, Amorim AF, Boustany A, Canales Ramirez C, Cardenas G, et al. Makaira nigricans. In: The IUCN Red List of Threatened Species 2011:

579 e.T170314A6743776 [Internet]. 2011 [cited 13 Jun 2020]. Available:
580 https://dx.doi.org/10.2305/IUCN.UK.2011-2.RLTS.T170314A6743776.en.
581 52. ICCAT. Report of the 2018 ICCAT Blue Marlin Stock Assessment Meeting.
582 SCRS/2018/008. International Commission for the Conservation of Atlantic Tunas
583 Collective Volume of Scientific Papers. 2018; 75: 813–888.
584 53. Sherley RB, Crawford RJM, Dyer BM, Kemper J, Makhado AB, Masotla M, et al. The
585 status and conservation of Cape Gannets Morus capensis. Ostrich. 2019; 90: 335–346.
586 doi:10.2989/00306525.2019.1684396
587 54. Distiller G, Altwegg R, Crawford RJM, Klages NTW, Barham B. Factors affecting adult
588 survival and inter-colony movement at the three South African colonies of Cape gannet.
589 Mar Ecol Prog Ser. 2012; 461: 245–255. doi:10.3354/meps09807
590 55. Crawford RJM, Sydeman WJ, Thompson SA, Sherley RB, Makhado AB. Food habits of
591 an endangered seabird indicate recent poor forage fish availability off western South
592 Africa. ICES J Mar Sci. 2019; 76: 1344–1352. doi:10.1093/icesjms/fsz081
593 56. BirdLife-International. Morus capensis. In: The IUCN Red List of Threatened Species
594 2018: e.T2269666 [Internet]. 2018 [cited 13 Jun 2020]. Available:
595 http://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22696668A132587992.en.
596 57. Ale SB, Mishra C. The snow leopard's questionable comeback. Science. 2018; 359: 1110–
597 1110. doi:10.1126/science.aas9893
598 58. Akçakaya HR, Keith DA, Burgman M, Butchart SHM, Hoffmann M, Regan HM, et al.
599 Inferring extinctions III: A cost-benefit framework for listing extinct species. Biol Conserv.
600 2017; 214: 336–342. doi:10.1016/j.biocon.2017.07.027
601 59. Winker H, Novellie P, Selier J, Birss C, Hraber H. Population trends and management

strategy tools for Cape Mountain Zebra. South African National Biodiversity Institute; 2016 pp. 1–46. Report Number: SANBI/BAM/STATS/2016/MZ/Final.

60. da Silva C, Winker H, Parker D, Kerwath SE. Assessment of smoothhound shark Mustelus mustelus in South Africa. Department of Agriculture, Forestry and Fisheries; 2019 pp. 1–29. Report No.: DEFF/FISHERIES/LSWG/AUG 2019.

61. Godfrey M, Godley B. Seeing past the red: flawed IUCN global listings for sea turtles. Endanger Species Res. 2008; 6: 155–159. doi:10.3354/esr00071

62. Boyd IL. Assessing the effectiveness of conservation measures: Resolving the "wicked" problem of the Steller sea lion. Biol Conserv. 2010; 143: 1664–1674. doi:10.1016/j.biocon.2010.04.006

Figures

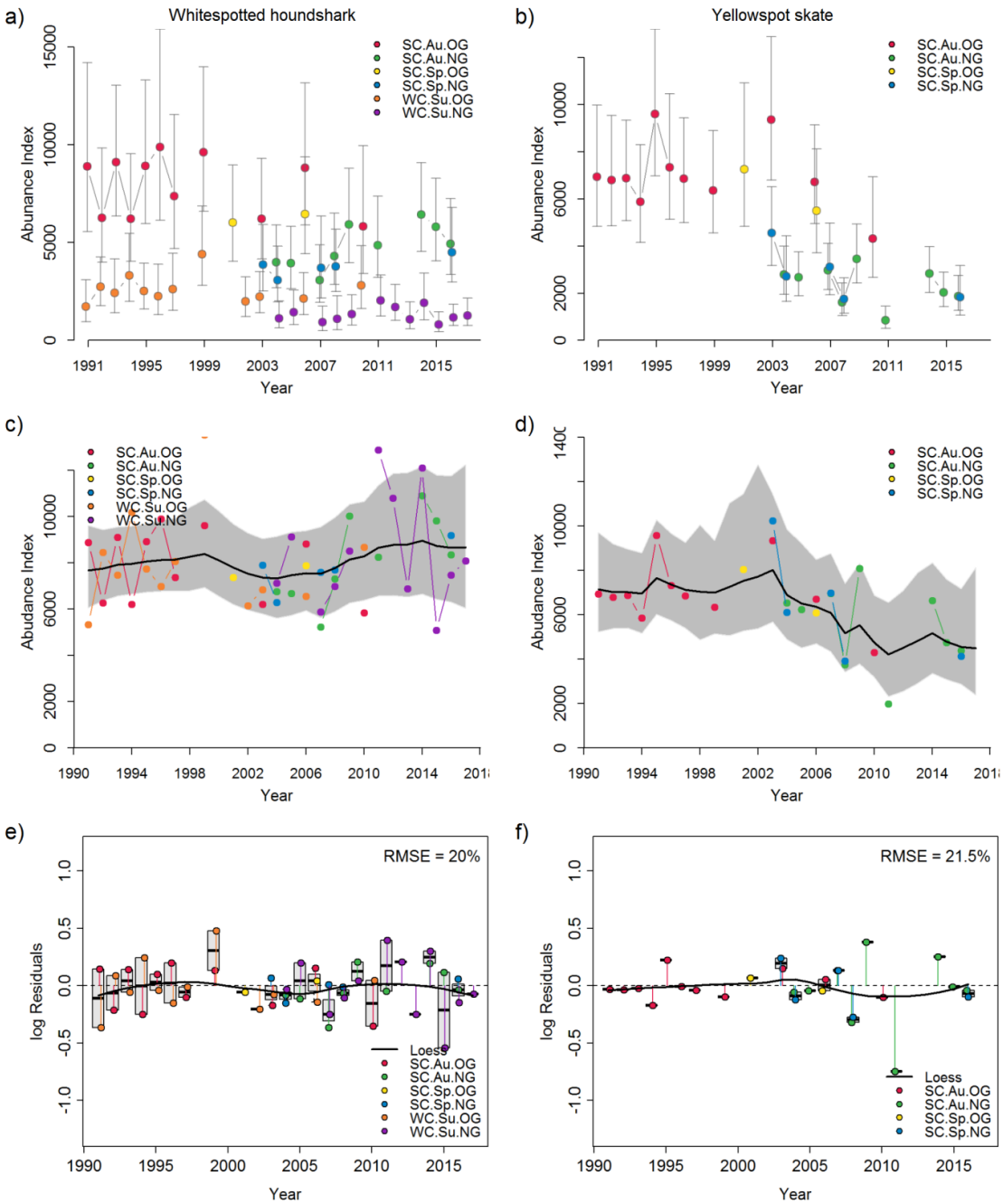


Figure 1: Input datasets of standardized relative abundance indices (± 95% CIs) from South African trawl surveys for two chondrichthyes species, (a) Whitespot Smoothhound *Mustelus palumbes* and (b) Yellowspotted Skate *Leucoraja wallacei* and Model diagnostics plots produced by JARA showing (c) – (d) 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.

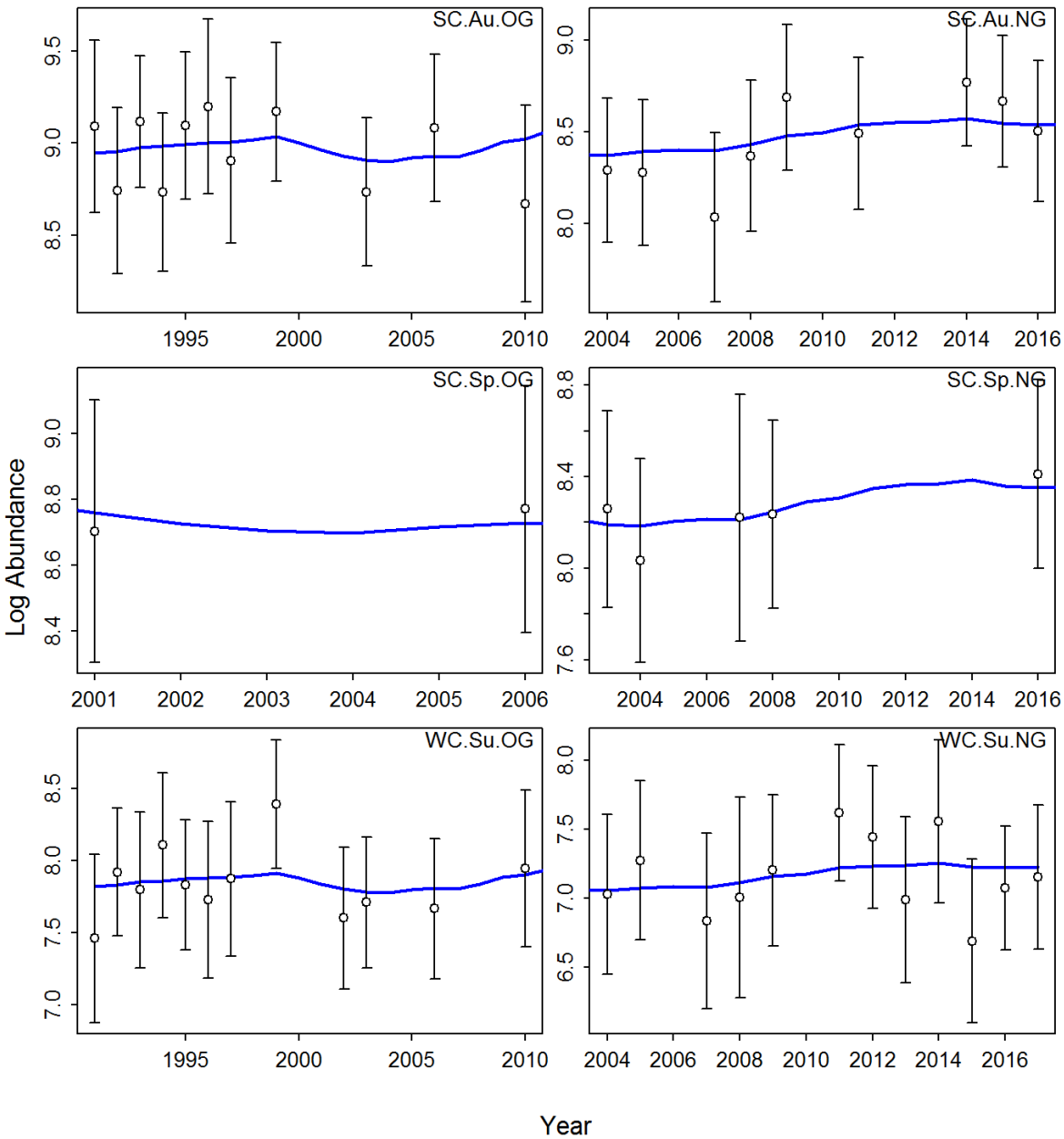


Figure 2: Model diagnostics plots produced by JARA using data from Whitespot Smoothhound *Mustelus palumbes*. Each panel shows the state-space model fit (blue line) to the natural logarithm (log) of the observed relative abundance data (white points), shown with 95% credible intervals (black lines) derived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

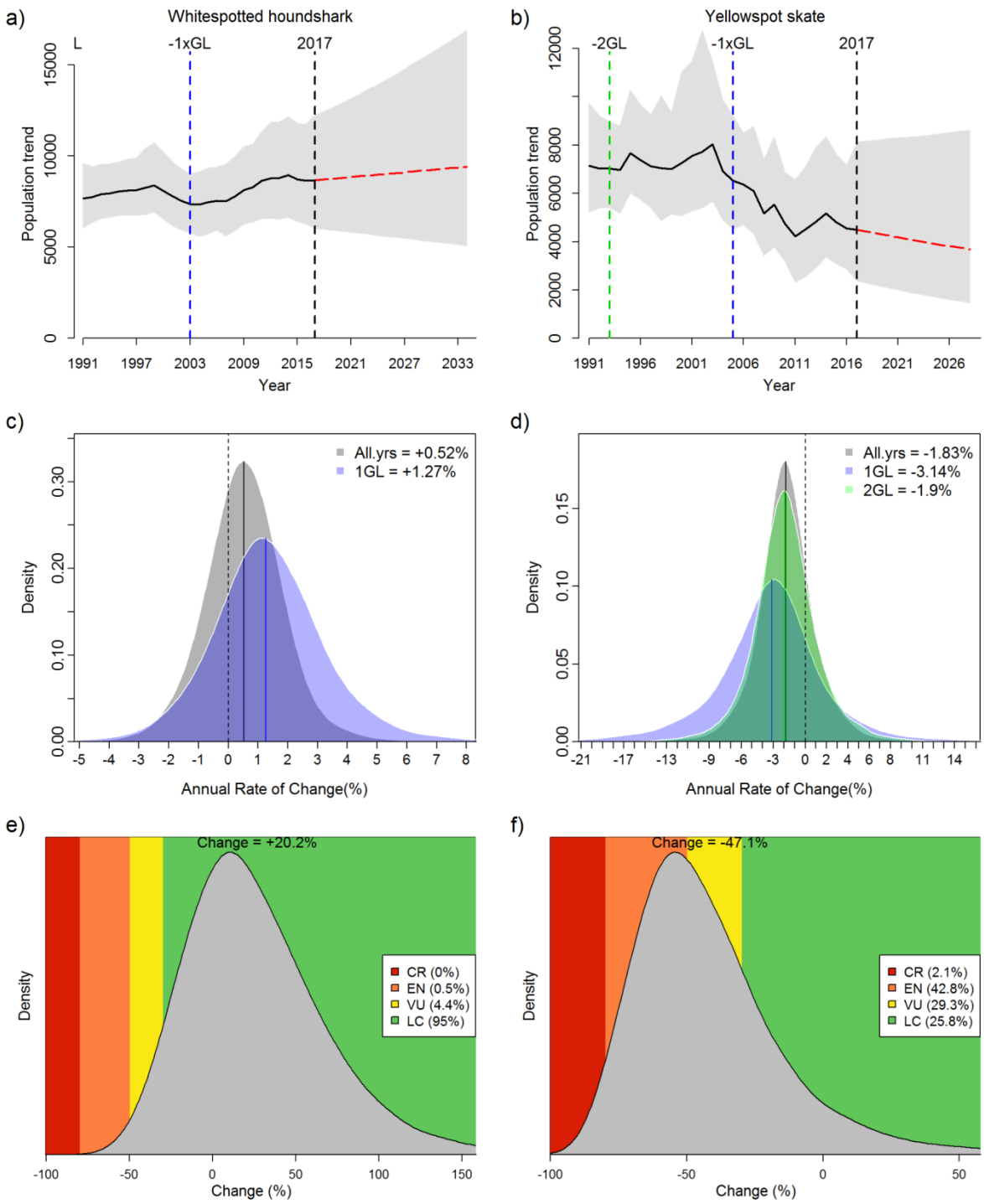


Figure 3: JARA decision-support plots for (left) Whitespot Smoothhound *Mustelus palumbes* and (right) Yellowspotted Skate *Leucoraja wallacei* showing (a) – (b) the overall JARA fit (black line) to the observed timeseries and the projected (red line) population trajectory over three generation length (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

636 medians (solid lines) shown relative to a stable population (% change = 0, black dashed line); (e) – (f) the
637 median change over three generation lengths ("Change = xx") and corresponding probabilities for rates of
638 population change falling within the IUCN Red List categories.

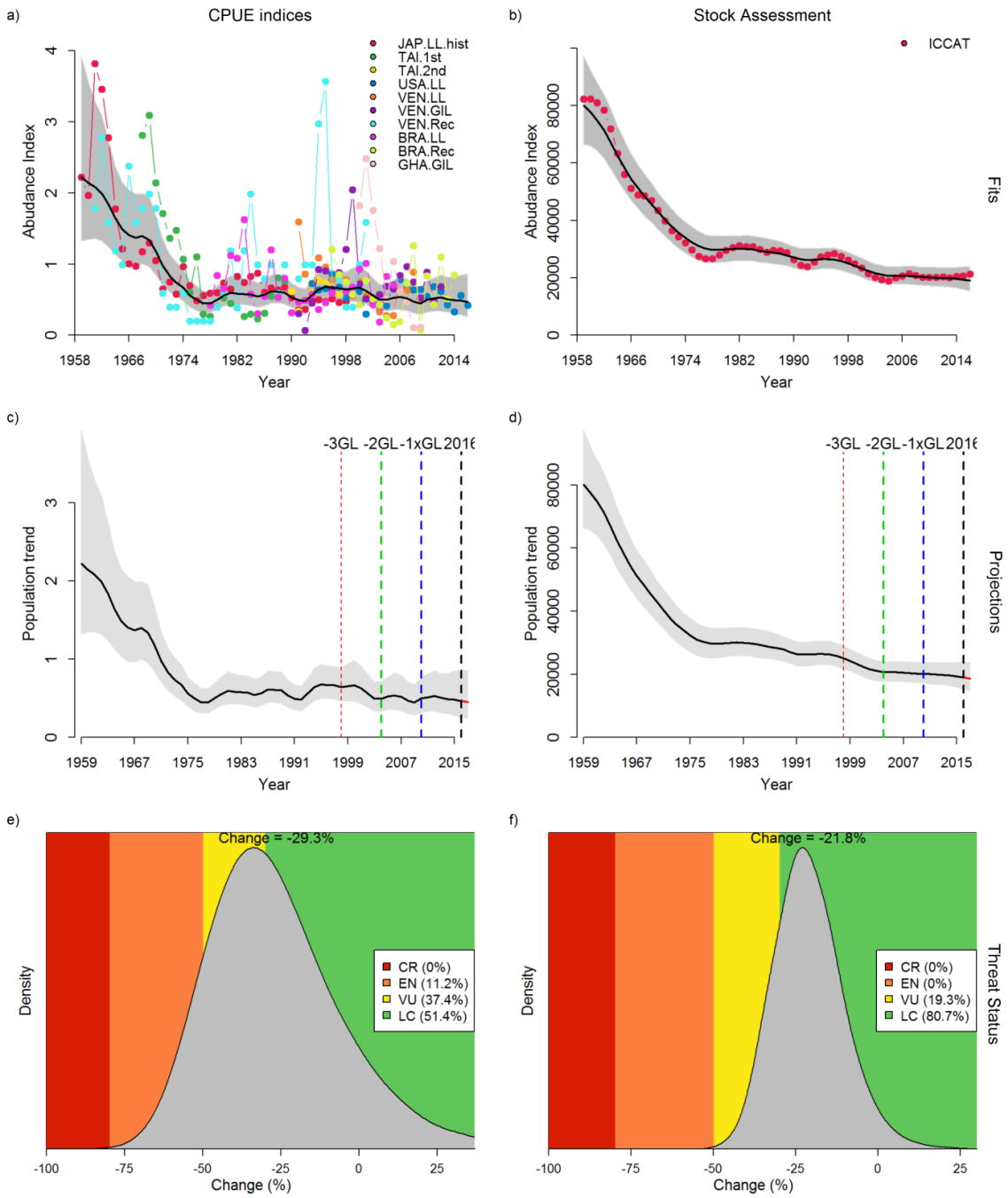


Figure 4: JARA plots for trend analyses of two types of abundance information for Atlantic Blue Marlin *Makaira nigricans*: (left) ten standardized catch per unit effort (CPUE) indices from multiple fishing fleets (1959–2016) and (right) estimated biomass estimates from the 2018 Atlantic stock assessment 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

647 timeseries with vertical dashed lines denoting the periods corresponding to 1–3 GL; (e) – (f) the median
648 change over three generation lengths ("Change = xx") and corresponding probabilities for rates of
649 population change falling within the IUCN Red List categories.
650

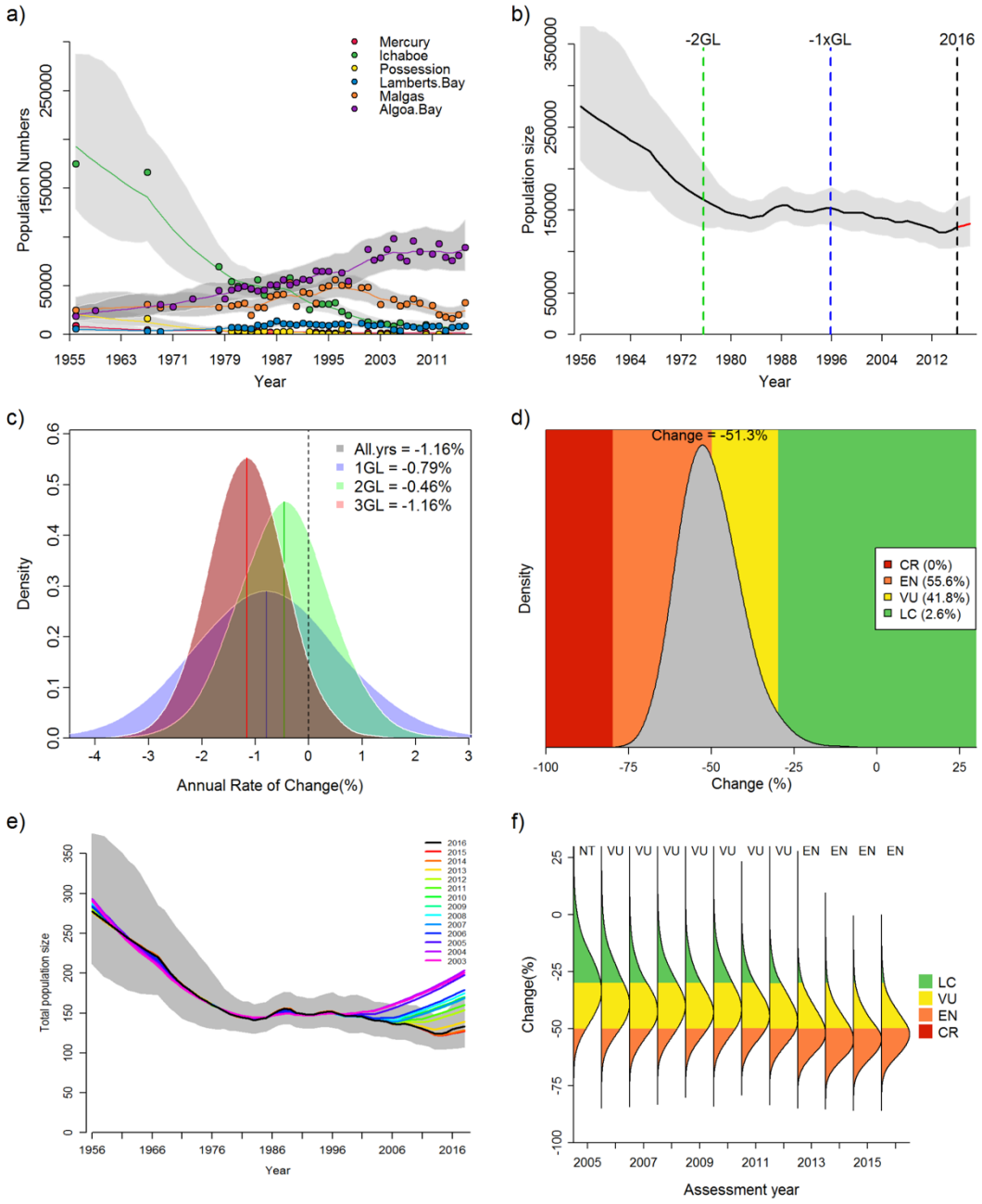


Figure 5: JARA decision-support plots for Cape Gannet *Morus capensis* showing (a) the JARA fits (coloured lines) and 95% credible intervals (grey polygons) to each observed abundance timeseries (coloured points); (b) the overall JARA fit (black line) to the observed timeseries and the projected (red 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.

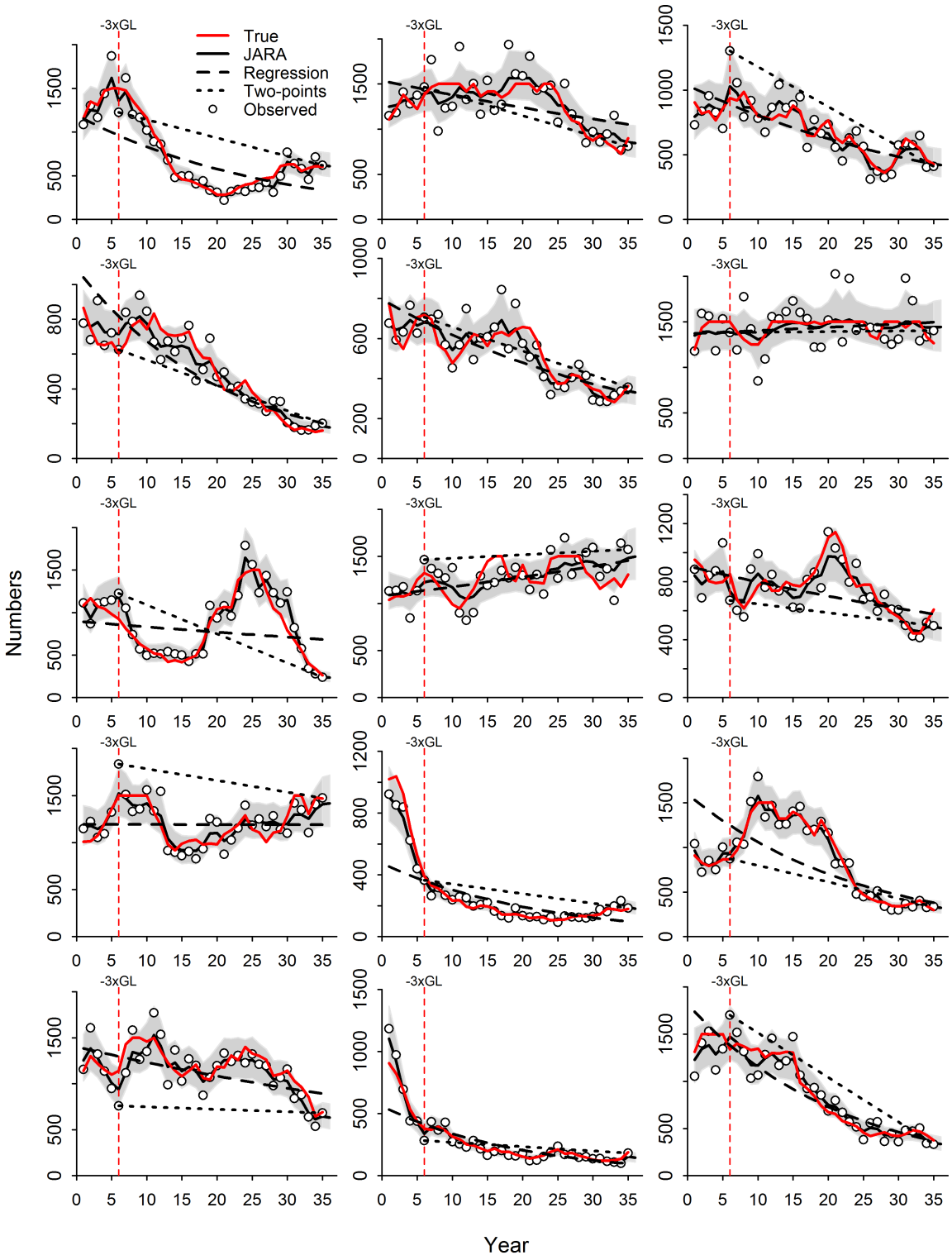


Figure. 6: Simulated 'true' abundance trajectories and observations for the first 15 of 200 simulation runs together with estimated abundance trends using three alternative estimation models (EMs): (1) 'Two-

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.

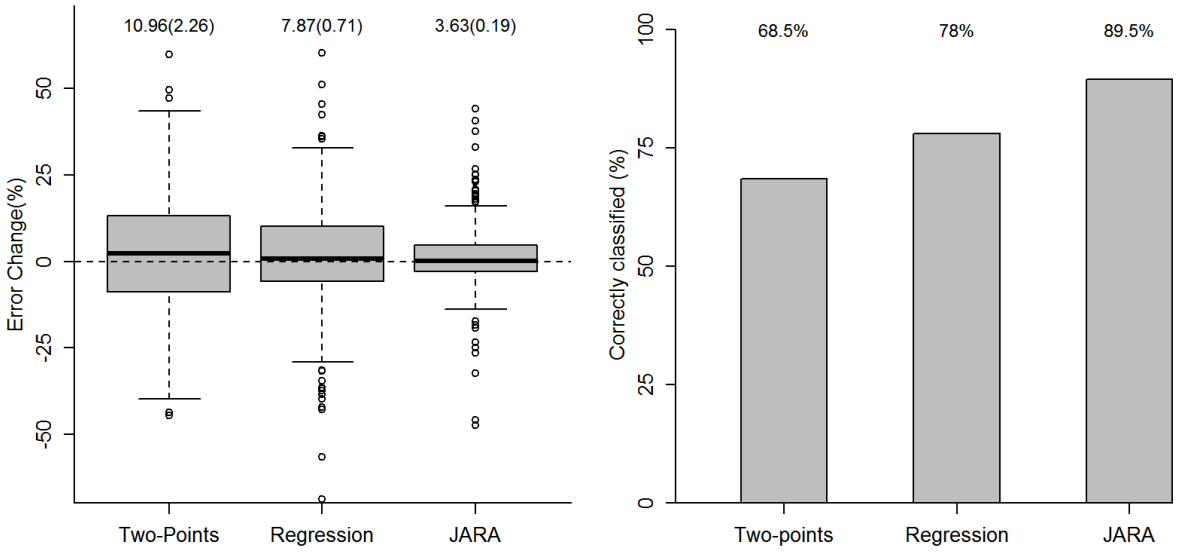
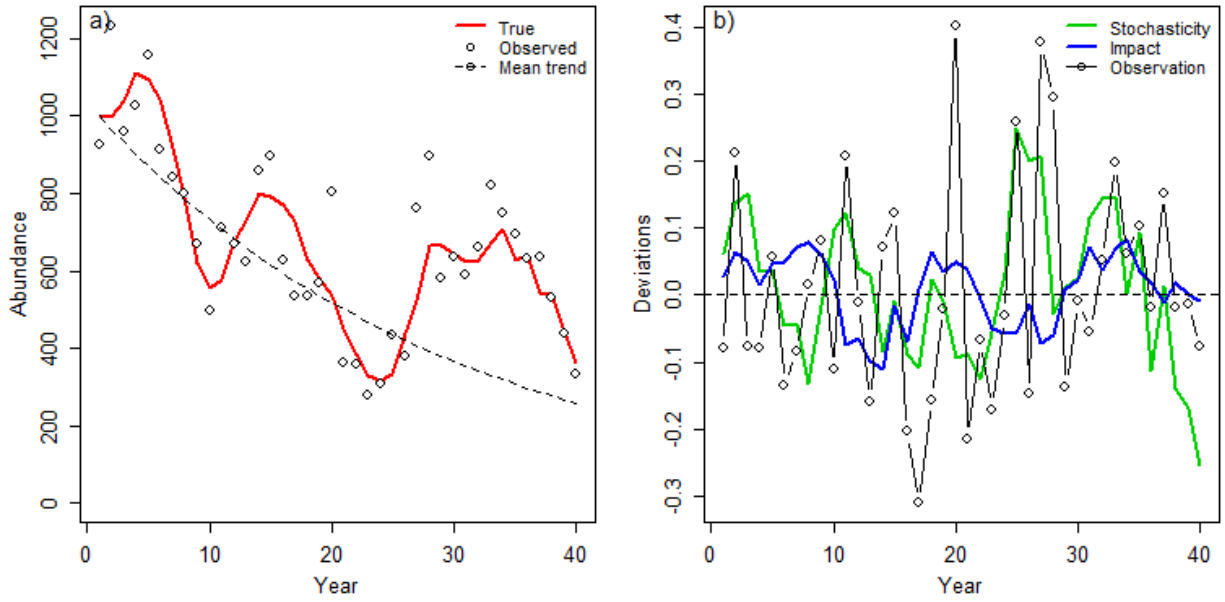


Figure. 7: Simulation testing results across 200 runs showing (a) boxplots of errors between the estimated and 'true' percentage change over 3 generation lengths (GL) with Median Absolute Errors and Median Errors (in brackets) displayed of each box and (b) barplots illustrating the percentages of correctly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

675 **Supplementary Information**

676



677

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

684

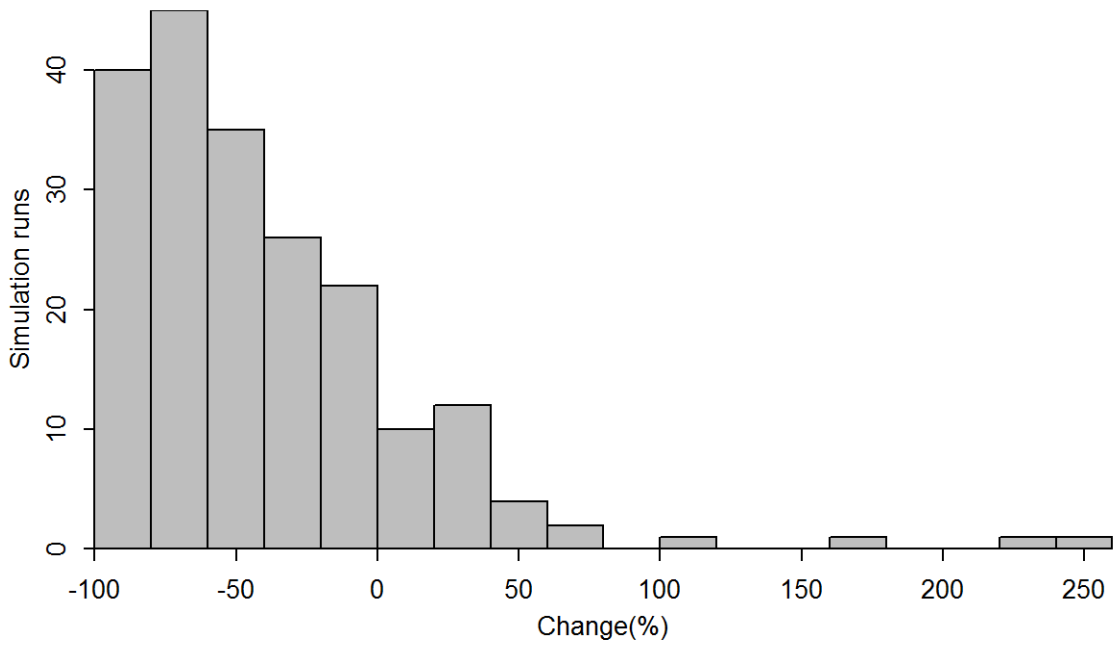


Figure S2: Simulated distribution of the 'true' %*C* generated from 200 simulation replicates of abundance trajectories over a 30-year assessment horizon.

```
#-----------------------------------------------------------------------------------------------------------
# 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()
```