ICESJMS-2021-196 (EA)

Reconciling conflicting survey indices of abundance prior to stock assessment[1]

Cassidy D. Peterson[1,2*], Dean L. Courtney[3], Enric Cortés[3], and Robert J. Latour[1]

[1] *Virginia Institute of Marine Science, William & Mary, P.O. Box 1346, Gloucester Point, VA 23062, USA*
[2] *Present address: Southeast Fisheries Science Center, National Marine Fisheries Service, 101 Pivers Island Road, Beaufort, NC 28516, USA*
[3] *Southeast Fisheries Science Center, National Marine Fisheries Service, 3500 Delwood Beach Rd, Panama City, FL 32408, USA*

[*] *Corresponding author: tel: +1 252 838 0885;* e-mail: cassidy.peterson@noaa.gov

Indices of relative abundance are one of the most important inputs into a stock assessment model. For many species, we must rely on several indices that routinely conflict with each other and which may result in biased and uncertain outputs. Here, we explored whether reconciled trends obtained from dynamic factor analysis (DFA) applied to conflicting indices can be used as a trend of relative abundance input into a stock assessment. We simulated an age-structured population of two coastal shark species in the southeast United States to generate multiple disagreeing indices, reconciled the indices using DFA, and then inserted both the multiple conflicting survey indices and the simplified DFA-predicted trend into respective stock assessment models. We compared the results of each stock assessment model to simulated values to evaluate the relative performance of each approach. We found that the DFA-based assessment generally performs similarly to the conflicting index-based assessment and may be a useful assessment tool in situations where conflicting indices with different selectivities, catchabilities, variances, and missing data are present. DFA assessment results were more consistent across simulation scenarios and outperformed many conflicting index assessments when surveys underwent shifts in catchability and the underlying stock abundance exhibited contrast.

Keywords: catchability, coastal shark, data conflict, dynamic factor analysis, relative abundance, stock synthesis.

[1] Note this manuscript details an extension study of Peterson, C. D., Wilberg, M. J., Cortés, E., and Latour, R. J. 2021. Dynamic factor analysis to reconcile conflicting survey indices of abundance. ICES Journal of Marine Science: 1-19.

Introduction

The purpose of a stock assessment is to assess the status of a given stock and provide management advice through the synthesis of multiple sources of information using quantitative models (Hilborn and Walters, 1992). When assessing the status of a stock, it is worth considering the relative importance of component datasets, particularly indices of relative abundance, since these are often considered to be key data inputs (Hinton and Maunder, 2003; Francis, 2011; Cortés *et al*., 2015). Indices of abundance serve to establish the trend in abundance for the resource over time and are often estimated from fishery-independent survey data. This trend is then scaled from relative to absolute abundance within the assessment using other information about the stock, including commercial and recreational catches.

Challenges associated with the biology and fishery dynamics of some species prevent logistically feasible and meaningful estimation of indices of abundance. Rare species or those with low catchabilities to particular gear types naturally have higher variability in survey catches, thereby increasing uncertainty in temporal patterns of the resource (Buckland *et al*., 2011). Particularly short- or long-lived species, which typically correspond to fast- and slow-growing species, respectively, require either more frequent sampling or longer time-series to adequately measure stock dynamics (Cortés, 2011). Species of low economic value are typically low priority, with few resources invested in survey programs and biological sampling regimes, which can result in data limitation (Stevens *et al*., 2000; Field *et al*., 2009; Cortés *et al*., 2015). Further, species with broad distributions can be challenging to monitor because of distributions that are too large to be reasonably assessed using a single survey program (Stevens *et al*., 2000).

For wide-ranging species that are not adequately sampled using a single survey, multiple spatially limited surveys are typically relied upon to gain a more complete interpretation of relative abundance patterns over time. Survey areas smaller than the distribution of the stock violate best practices by not surveying the entire population (Hilborn and Walters, 1992). Different surveys may utilize different gears or operational protocols, sample in different areas or during different times of the year, which result in a separate proportion of the stock being available to the unique sampling gear (Maunder *et al*., 2006; Cook, 2010; Ono *et al.*, 2018). When the entire population is not available to the gear, the assumption that survey catches are proportional to the total stock abundance may not hold true (Maunder *et al*., 2006; Wilberg *et al*., 2010; Maunder and Piner, 2017). Since each survey is not necessarily measuring the same signal in the population, and survey programs experience high levels of uncertainty (process and observation error; Maunder and Piner, 2017), it is common for the temporal trends of multiple indices of abundance to be in disagreement with each other (Schnute and Hilborn, 1993; Conn, 2010a). Conflicting trends in relative abundance lead to uncertainty in the status of the resource and convey contradictory information to a stock assessment model (Schnute and Hilborn, 1993; Conn, 2010a; Francis, 2011).

The sandbar shark (*Carcharhinus plumbeus*) is an example of a species with a large spatial distribution (Springer, 1960; Heist *et al.*, 1995) which also grows slowly (Sminkey and Musick, 1995; Brewster-Geisz and Miller, 2000; Baremore and Hale, 2012), experiences low catches, exhibits clear migratory patterns, spatially segregates, and is of relatively low economic value (Bigelow and Schroeder, 1948; Springer, 1960; Kohler *et al*., 1998). In the most recent stock assessment (SEDAR, 2017), 11 indices of abundance were included, and an additional three indices were considered in the prior assessment (SEDAR, 2011). Unsurprisingly, the 11 relative abundance indices showed conflicting patterns over time. Using a hierarchical clustering approach, the indices were split into two groups: those that generally increased (‘Pos’ indices),

and those that decreased ('Neg' indices; Courtney, 2017). Each grouping of indices ('Pos' and 'Neg') was separately introduced to the sandbar shark assessment model as a unique sensitivity run in an effort to characterize various potential states of nature, a largely supported approach to dealing with data conflict (Schnute and Hilborn, 1993; Francis, 2011). The base runs with all conflicting indices, the 'Pos' sensitivity run and the 'Neg' sensitivity run, each resulted in a different stock status determination; hence, collectively, results obscured the understanding of resource status.

In contrast to large coastal sharks, small coastal sharks, like the Atlantic sharpnose shark (*Rhizoprionodon terraenovae*), generally grow faster and exhibit increased genetic stock structure between the Gulf of Mexico and southeastern US Atlantic due to comparatively lower movement rates (Loefer and Sedberry, 2003; Davis *et al*., 2019). Despite decreased home ranges, small coastal sharks are still not comprehensively sampled by a single or few surveys. For example, SEDAR 34 (2013) used 15 indices of abundance in the base run of the combined region assessment for Atlantic sharpnose shark.

Among other methods, dynamic factor analysis (DFA) has been explored as an approach to reconcile conflicting survey indices (Zuur *et al*., 2003a; Azevedo *et al*., 2008; Peterson *et al*., 2017, 2021). Dynamic factor analysis is a state-space, multivariate, dimension reduction approach (Zuur *et al*., 2003a, 2003b). Notably, Azevedo *et al*. (2008) used DFA to reconcile conflicting patterns among abundance indices of two species of Iberian anglerfish. The DFA-predicted common trends for the white anglerfish (*Lophius piscatorius*) were then used as inputs into a biomass dynamic assessment model, resulting in a better fitted model as measured by residual mean square error, narrower confidence intervals, and lower bias in parameter estimates as measured by residual bootstrapping. The authors suggested that DFA could also be explored in more complex assessment modeling frameworks (Azevedo *et al*., 2008).

We sought to determine if there is validity to reconciling conflicting indices of abundance using DFA prior to fitting an assessment model vs. inputting multiple pieces of contradictory relative abundance information into an assessment. Despite general advice to manipulate data as little as possible prior to fitting an integrated assessment model (Maunder, 2001; Maunder and Punt, 2013; Methot and Wetzel, 2013), the incorporation of conflicting abundance indices within an integrated assessment model results in data conflict within the model likelihood (Carvalho *et al*., 2021). Current *ad hoc* approaches to address data conflict within integrated assessment models include removing or downweighting conflicting data within the model likelihood, which can affect both parameter estimation and the resulting management advice obtained from an integrated assessment model. In contrast, the results from this study provide guidance regarding the effectiveness of reconciling conflicting indices of abundance before inclusion into a stock assessment model, which may serve as both a useful diagnostic for evaluating the effects of data conflict on parameter estimation and the resulting management advice within the assessment framework and provide an objective approach for addressing data conflict when there are multiple conflicting indices of abundance in the same stock assessment model.

We chose two sharks (sandbar shark and Atlantic sharpnose shark) as representative large and small coastal shark species, respectively, in our study. Each species was chosen based on availability of recent stock assessments. Each stock was simulated using an age-structured model, including generation of conflicting indices of abundance (Peterson *et al.*, 2021). Conflicting indices were input into a stock assessment model ('CI assessment') and results were compared to those generated from inputting a reconciled DFA trend into an equivalent stock assessment model ('DFA assessment').

Methods

We evaluated stock assessment performance when including multiple conflicting survey indices (CI assessment) compared to a dimension-reduced DFA trend (DFA assessment) as the relative abundance information for the Atlantic sharpnose shark and the sandbar shark (Figure 1). Simulation sensitivities included (1) changes in the underlying temporal pattern of population abundance, (2) generation of conflicting survey indices through time-varying catchability, and (3) missing years of survey information. Unaccounted-for shifts in catchability (q) were induced to generate conflicting indices of abundance following Peterson *et al.* (2021; Figure 1).

Two age-structured operating models (OMs) were constructed with characteristics unique to each species. Corresponding estimating models (EMs) were developed using Stock Synthesis (version 3.24; adapted from SEDAR 54, SEDAR, 2017), an integrated stock assessment modeling framework. The simulated conflicting survey indices were input into the EMs as the relative abundance indicators for the CI assessment, while the DFA trend was input into the EMs for the DFA assessment.

Study species

Atlantic sharpnose sharks, though historically assessed as a single stock off the US Atlantic coast and in the Gulf of Mexico (SEDAR, 2013), have recently been found to exhibit stock structure between the two areas (Davis *et al*., 2019). The Atlantic stock was simulated in the current study. In the US Atlantic, female Atlantic sharpnose sharks have a median age at maturity of 1.6 years (Loefer and Sedberry, 2003), maximum longevity of 23 years (Frazier *et al*., 2014), an annual reproductive cycle, and an average litter size of 4–5 pups (Castro, 2009). Atlantic sharpnose sharks are moderately productive, with steepness values estimated at 0.56 (SEDAR, 2013). Because SEDAR 34 (SEDAR, 2013) assessed the stock using a single-sex, state-space, age-structured production model, we only had sufficiently available data to generate a single-sex simulation in our study.

Sandbar sharks comprise a single genetic stock in the US Atlantic and Gulf of Mexico (Heist *et al*., 1995) and were most recently assessed using Stock Synthesis (SEDAR, 2017). Female sandbar sharks have a median age at maturity of 14 years (Baremore and Hale, 2012), estimated longevity of 31 years (SEDAR, 2017), a biennial or triennial reproductive cycle (which is, therefore, modeled as 2.5 years; Baremore and Hale, 2012; SEDAR, 2017), and an average litter size of eight pups (Baremore and Hale, 2012). Sandbar sharks have particularly low productivity, with steepness estimated at 0.3 (SEDAR, 2017). Since SEDAR 54 (SEDAR, 2017) employed a two-sex assessment model, we had sufficient information available to build a two-sex sandbar shark simulation in our study.

Operating model

The operating model (OM) was based on an age-structured model (see Table 1 for OM equations and Tables S1–S4 in the Supplementary material for simulated parameter values), with a low-fecundity stock–recruitment relationship (LFSR; Taylor *et al*., 2013). The Atlantic sharpnose shark OM had one fishing fleet and either three or four surveys, while the sandbar shark OM had four fishing fleets and seven surveys. Modeled fishery and survey selectivities were based on those estimated in recent assessments (Figure 2; SEDAR, 2013; SEDAR, 2017). We assumed no discarding and no spatial structure for either species. These decisions were made to mirror the

limited data availability and the resulting lack of spatial structure within models to assess these species in practice.

Numbers of fish of sex s at age a in each year y ($N_{s,a,y}$) was a function of those that survived from the previous age and year, removing those that died from sex- and age-specific annual total mortality ($Z_{s,a,y}$; Table 1, equation 1). Recruits were defined as those individuals that survived their first year of life, where β, Z_{min}, and Z_0 are parameters of the LFSR function (Taylor *et al.*, 2013). The spawning output, measured in number of neonates birthed each year ($Npups_y$), was the product of the number of females-at-age in a given year ($N_{s=female,a,y}$) divided by reproductive periodicity for that species (RP; equation 2), female maturity-at-age ($p_{s=female,a}$), and fecundity-at-age (f_a) summed over ages. We assumed a 1:1 sex ratio at birth. Sex- and fleet-specific yearly catches-at-age ($C_{s,a,y,i}$) were calculated using sex-, age-, year-, and fleet-specific fishing mortality ($F_{s,a,y,i}$). Total fishing mortality at sex, age, and year ($F_{s,a,y}$) was converted to fleet-specific fishing mortality at sex, age, and year by multiplying $F_{s,a,y}$ by fleet-specific selectivity ($sel_{a,s,i}$) and the proportion of fishing mortality attributed to fleet i (δ_i; equation 3). Total mortality for each sex, age, and year ($Z_{s,a,y}$) was calculated by summing $F_{s,a,y,i}$ across fleets and adding natural mortality-at-sex and -age (equation 4). Indices of abundance indexed by sex, age, year, and survey j ($I_{s,a,y,j}$) were generated by multiplying $N_{s,a,y}$ by the survey-specific annual catchability ($q_{y,j}$) and vulnerability (i.e. gear selectivity) for each sex, age, and survey ($v_{s,a,j}$) adjusted by lognormal error ($\varepsilon_{s,a,y,j}$), where $\sigma_{s,a,y,j}$ is the lognormal standard deviation of the yearly sex- and age-structured index for each survey j. Note that we implemented survey variability by defining survey-specific coefficients of variation (CV_j). Thus, we multiplied the CV_j by the expected value of the survey index (following $CV \times \mu = \sigma$) to generate $\sigma_{s,a,y,j}$ (equation 5), which was consequently indexed by sex, age, year, and fleet. Yearly survey-specific indices ($I_{y,j}$) were calculated by summing $I_{s,a,y,j}$ across age and sex (equation 5). Lastly, we assumed that ageing was not conducted each year to mirror the limited data availability for these species in practice. Ages generated from fishery-dependent and fishery-independent sources were converted to lengths for each individual, n, using the von Bertalanffy growth equation (equation 6). Ages were jittered by 20% as an additional source of observation uncertainty. The number of length composition observations taken varied based on the survey and species; more length observations were taken for the Atlantic sharpnose shark (~90–145 observations year^{-1}) compared to the sandbar shark (~1–50 observations year^{-1}). Theoretical age at length 0, t_0, was assumed to be constant, and asymptotic maximum length, L_∞, and the growth coefficient, K, were implemented stochastically for each individual and with correlation of $\rho = -0.9$ (Bertsekas and Tsitsiklis, 2002).

The operating model for each species used estimated parameters from the most recent assessments (SEDAR, 2013; SEDAR, 2017) where applicable (Tables S1–S4 in the Supplementary material). Operating model simulations were conducted in the programming language R (version 3.6.2; R Core Team, 2019). Representative R code is available via https://github.com/cassidydpeterson/DFA_Simulation_and_Assessment.

Dynamic factor analysis

Survey indices generated in the OMs were inputted into a dynamic factor analysis (DFA) model (Figure 1). Dynamic factor analysis is a multivariate, dimension-reduction, state-space approach designed for nonstationary time-series (Zuur *et al.*, 2003a; Holmes *et al.*, 2014). The form of the DFA model follows equation 7 in Table 1, where y_t is a vector of n log-transformed then standardized time-series at each time t, α_t is a vector of m common trends ($m < n$) with associated

error (η_t), $\boldsymbol{\Gamma}$ is an $n \times m$ matrix containing factor loadings, which indicate the influence of the common trend on each input time-series, and ε_t and η_t represent observation and process error terms, assuming a multivariate normal distribution with mean **0** and covariance matrices $\boldsymbol{H}$ and $\boldsymbol{Q}$, respectively. $\boldsymbol{Q}$ was set equal to the identity matrix (Zuur *et al*., 2003a, 2003b; Holmes *et al.*, 2014). Survey uncertainty was propagated through DFA analyses by setting average survey *CV* as the diagonal elements of the $\boldsymbol{H}$ matrix (variance) with no covariance between surveys. Survey indices with lower *CV*s were generally weighted more heavily within the DFA, as denoted by larger factor loadings (Peterson *et al*., 2021).

We assumed $m = 1$ to only generate a single DFA trend, and we excluded explanatory variables. DFA modeling assumptions include normality, independent error, and homogeneity of residuals (Zuur *et al*., 2003a). We applied an alternative rescaling approach to the log-transformed indices (comparable to z-scoring) prior to running the DFA, which allowed for backtransformation of the resulting DFA trend out of log-space (see Supplementary material; Peterson *et al*., 2021). Models were fitted using the state-space multivariate autoregressive modeling package 'MARSS' in R (Holmes *et al*., 2014).

Simulation scenarios

We explored the results of assessments conducted using multiple indices of abundance as survey index inputs compared to those generated using a single, DFA-predicted trend as a survey index input across various scenarios. The scenarios were intended to approximate various realistic conditions that may violate assumptions of index of abundance generation to explore how the resulting assessments perform under each different trial. Simulation scenarios explored in the current study were the same as those explored in Peterson *et al*. (2021; Tables 2–3).

In the Atlantic sharpnose shark simulation, we explored DFA performance across: (1) underlying population abundance patterns by varying fishing mortality, and (2) inducing survey indices to conflict by shifting catchability (both as a knife-edged shift or a gradual shift). We also explored the effects of survey variability by altering survey-specific *CV*s and adding a partial, fourth survey index (see Supplementary material). In the sandbar shark simulation, we explored DFA performance across: (1) conflicting survey indices as generated by implementing changes in patterns of catchability over time, and (2) missing years of survey data vs. full survey data.

Each simulation scenario was replicated 100 times. Simulations spanned 65 or 90 years, with survey data collected beginning after 40 or 51 years in 1993 and 1979 for Atlantic sharpnose and sandbar sharks, respectively (Tables 2–3; Figure 3), to mirror realistic data availability. For Atlantic sharpnose shark, the knife-edged shift in catchability began in 2003, and the gradual change followed a linear ramp from 2003 to 2013 or 2018, with catchability shifting from 0.00025–0.001 (q_1) or 0.003–0.004 (q_3). For the sandbar shark, catchability increased or decreased to/from 0.01 to 0.045 in a linear or stepwise pattern over years 1979 to 2018 (Figure 3).

Estimating model

We generated two EMs for each simulation scenario: (1) a conflicting index (CI) EM in which multiple simulated survey indices were included, and (2) a DFA EM in which a single DFA-predicted trend was input in the EM as the measure of relative abundance. All equations used in the Stock Synthesis assessment platform can be found in Methot and Wetzel (2013). Stock Synthesis is an integrated analysis framework which can incorporate several sources of data, accommodate complex model configurations, scale between data-limited and data-rich

assessments, and propagate uncertainty (Methot and Wetzel, 2013). Stock Synthesis is a flexible modeling platform that has been shown, under simplifying assumptions, to provide useful management advice in data-limited situations (Wetzel and Punt, 2011; Cope, 2013). Analyses were conducted in R (version 3.6.2; R Core Team, 2019) and Stock Synthesis (Methot and Wetzel, 2013), using the R package r4ss (Taylor *et al*., 2019, 2021) for communication between Stock Synthesis and R.

Estimating model assumptions

Following each simulation, the Atlantic sharpnose shark stock was assessed through a female-only model, and the sandbar shark stock was assessed with a two-sex model, each with one area and no discards. Surveys were assumed to take place in the middle of the year, and catches-at-age were inferred from the fishery length composition data. Maximum age was assumed to be 27 years (Frazier *et al*., 2014) and 31 years for the Atlantic sharpnose and sandbar shark, respectively, and a plus group was implemented for ages greater than or equal to 18 years in the Atlantic sharpnose shark model. Based on the length of simulated survey indices, the starting year of each iteration was chosen as 1993 for the Atlantic sharpnose shark and 1979 for the sandbar shark, resulting in an end year of 2018 for each species.

Estimating model data

Age- and sex-specific instantaneous natural mortality was fixed except for pre-recruits (age 0) which was defined by the LFSR function. Fecundity was implemented through a time-invariant fecundity-at-age vector. The first mature age was assumed to be two years for Atlantic sharpnose sharks and 13 years for sandbar sharks. Von Bertalanffy growth was implemented following Schnute's (1981) three-parameter reparameterization (Methot and Wetzel, 2013). Growth parameters were independently calculated and fixed within the EM model. Weight-at-length was assumed to follow the allometric growth equation, where parameters were fixed based on values independently estimated in SEDAR (2013, 2017). Following SEDAR 54 (SEDAR, 2017), all life history parameters were fixed at their simulated values in the EM for both species.

All data included in the stock assessment model were generated from the OM. The index of abundance error was assumed to follow a Student's *t*-distribution in log space with 30° of freedom, which approximates a lognormal distribution (Methot, 2013). Indices of abundance were rescaled prior to DFA model fitting, as presented in Peterson *et al.* (2021; see Supplementary material), and the resulting backtransformed DFA trend was in arithmetic space with lognormal error. The backtransformed DFA trend was input into the EM with associated lognormal error.

For the DFA EMs, the resulting DFA trend was, by definition, representative of all length observations from each underpinning survey. Therefore, we combined length observations from each survey by taking a weighted average, where length observations (number of fish observed in each length bin in each survey per year) were weighted by relative size of the corresponding factor loading ($\Gamma_j \div \max(\boldsymbol{\Gamma}) \; \forall \; j$; see Supplementary material for additional details on treatment of composition data). Recall that factor loadings are elements of the $\boldsymbol{\Gamma}$ matrix and indicate the relative strength of the effect of the resulting DFA index on each survey. We assumed that if the DFA index was not heavily influenced by a survey, then the length observations of that survey should have less weight in the stock assessment. Age compositions were not included in the EMs.

Estimating model stock–recruit relationship

The survival-based stock–recruitment function (LFSR; Taylor *et al.*, 2013) was implemented within the EMs, which requires specification of three parameters: $ln(R_0)$, β, and s_{frac}. $ln(R_0)$ is the natural logarithm of virgin recruitment (where recruits were defined as age 0) and was estimated with no prior. The parameter β controls the shape of the density-dependent survival curve of prerecruited pups, and s_{frac} defines the fraction of pup survival when the population approaches zero. Both β and s_{frac} were fixed due to estimation concerns associated with assessing a 'one-way trip' (Hilborn and Walters, 1992). The standard deviation of recruitment (σ_R) was fixed at 0.5. Adjustments for recruitment variability bias were made following Methot and Taylor (2011). We assumed no environmental link or autocorrelation in recruitment.

Estimating model selectivity

Following the age-structured simulation models, we assumed that fishery and survey selectivities were age-structured. Simulated selectivity curves were based on observed fisheries and surveys for each species. All selectivity parameters were estimated using normal priors. EM selectivity curves closely approximated simulated selectivity-at-age (Figure 2).

In the Atlantic sharpnose shark EM, the fishing fleet was dome-shaped and modeled using a double normal selectivity function. In the CI EM, each survey was assumed to follow logistic selectivity corresponding to the OM simulations. In the DFA EM, the DFA-predicted index was assumed to follow a logistic selectivity function with time-blocks when survey indices contained missing data (Table S5 in the Supplementary material). All selectivity parameters were estimated with informative priors in the Atlantic sharpnose shark EM.

In the sandbar shark EM, fishing fleets were assumed to follow dome-shaped, double normal selectivity (fleets 1 and 3) or logistic selectivity (fleets 2 and 4). When age-specific selectivity varied by sex (fleets 1 and 3), a male offset was implemented to account for sex-specific differences in selectivity. In the CI EM, dome-shaped, double normal selectivities with sex-specific offsets were implemented for all surveys, with the exception of one (survey 2) which assumed logistic selectivity with no male offset (see Table S5 in the Supplementary material for more information on how selectivity was specified in the EM). In the DFA EM, where the DFA-predicted index was implemented, selectivity-at-age was estimated as a random walk (Figure 4). Random walk selectivity curves were not sex-specific.

When DFA was fitted to surveys with missing years of data (Atlantic sharpnose scenarios with four surveys; sandbar missing data scenarios), time-blocks were implemented to account for changes in selectivity when not all surveys were conducted (Table S5 in the Supplementary material). For example, if three surveys A, B, C were conducted, and survey A ran from years 1 to 10, survey B ran from years 4 to 8, and survey C ran from years 6 to 10, four time-blocks would be created: time-block 1 for years 1–3 where only survey A was run, time-block 2 for years 4–5 when surveys A and B were run, time-block 3 for years 6–8 when surveys A, B, and C were run, and time-block 4 from years 9–10 when surveys A and C were run. This way, there was a unique selectivity pattern for each combination of surveys. An age-based, random walk selectivity function was used to approximate the resulting selectivity curve. Starting values for each time-block represented an average of the selectivities for each survey that were fitted using Stock Synthesis selectivity helper excel spreadsheets (available via https://vlab.ncep.noaa.gov/web/stock-synthesis/document-library, accessed June 2020). The parameters of the resulting random walk selectivity function were used as the initial values for selectivity within the EM.

Estimating model post-run data weighting
We chose to implement the Francis (2011) weighting technique for length composition data. Essentially, after the EMs were run, input sample sizes of length composition observations were iteratively "right-weighted" to account for the correlated length observations obtained from catches and survey observations by reducing the observed sample size of length observations to the effective sample size. The effect of right-weighting length composition data consequently gives greater weight to indices of abundance (Courtney *et al*., 2017a; Francis, 2017). No additional variance was added to the survey index variances.

Estimating model convergence
We could not perform a full set of diagnostic procedures (e.g. Carvalho *et al*., 2021) for the EMs fitted to each iteration of every OM simulation scenario because of the large number of model runs. Consequently, a limited number of model-fit diagnostics commonly applied in data-moderate, age-structured stock assessments for sharks (e.g. Courtney, 2016; Courtney *et al*., 2017b) were evaluated here for the first iteration of each EM fitted to an OM scenario. These included a visual comparison of the predicted vs. simulated stock abundance, predicted vs. observed indices of abundance, annual length composition, and the aggregated annual length composition by fleet (Figures 5–6). These also included visual inspections of predicted recruitment relative to the assumed stock–recruitment relationship, as well as the residuals from the predicted vs. observed indices of abundance, annual length compositions, and estimated annual recruitment deviations for unexpected patterns or trends. Model-fit diagnostics for subsequent iterations were assumed to follow those of the first iteration. In contrast, diagnostics for the remaining iterations included visual comparison of the EM predicted vs. the OM simulated abundance (in numbers). EM fits that did not pass these model-convergence diagnostics or that failed to converge were rerun with alternate starting values or were excluded from subsequent analyses if they failed the convergence diagnostics over multiple iterations.

Estimating model performance
Assessment performance was inferred from accuracy of estimated fishery parameters relevant for management purposes, namely depletion (current stock size / virgin stock size in numbers) and fishing mortality in the final year of the simulation (or final year of fishing if fishing mortality was set at zero in the final years of the simulation; F_{final}) following Courtney *et al*. (2016). Estimated quantities are presented as relative values (i.e. estimated value / simulated value). See Supplementary materials for estimated virgin abundance (N_0 in numbers), root-mean square error (RMSE) in estimated abundance in numbers, and relevant maximum sustainable yield-based reference points.

Results
In Peterson *et al*. (2021), we described the performance of the DFA model across simulation scenarios; in the current study, we describe the performance of the assessment models when DFA-predicted indices are implemented in a stock assessment (DFA assessment) compared to the standard practice of including multiple survey-based indices of abundance (CI assessment). We evaluated CI vs. DFA assessment performance for the Atlantic sharpnose shark and the sandbar shark across (1) changes in the underlying temporal pattern of population abundance, (2) generation of conflicting survey indices through time-varying catchability, and (3) missing years

of survey information. (See Supplementary materials for details on how CI and DFA assessments performed across survey variability, number of surveys, and with respect to management reference points.) Recall that assessment model performance was measured by the accuracy of estimated parameters.

Model fitting

When simulations were run with no error, the EMs were generally capable of accurately replicating the OM dynamics. All Atlantic sharpnose shark EMs converged, and fewer CI assessments converged (94.8%) than DFA assessments (99.9%) in sandbar shark EMs.

Underlying pattern in population abundance

Under the 'no change in underlying population abundance over time' (ConstF) Atlantic sharpnose shark scenario, assessment model performance was generally poor when surveys experienced time-varying catchability (Figures 7–8). The CI assessment was generally more accurately and precisely capable of estimating relative depletion and terminal fishing mortality than the DFA assessment when surveys experienced unaccounted-for time-varying catchability.

When the underlying population decreased in a one-way trip (Inc F), the CI assessment more accurately estimated relative depletion and F_{final} for constant and knife-edge catchability patterns, but not when catchability underwent a gradual shift (Figures 7–8). When the population increased in a one-way trip (Dec F), the DFA and CI assessment results were similar, though the DFA assessment more consistently and accurately estimated depletion and F_{final}, especially when four surveys were used (Figures 7–8).

The outcome for the fishing mortality scenario that resulted in a population that underwent a decrease then increase in size (U F) seemed to indicate that both the CI and DFA assessment models performed fairly well and neither assessment performed better in all scenarios (Figures 7–8 and Figures S2–S5 in the Supplementary material). In almost all scenarios in which time-varying catchability was simulated along with a U-shaped fishing mortality pattern, the DFA assessment results were more precise than those from the corresponding CI assessment (as denoted by shorter violins), except in the scenarios in which the surveys underwent a knife-edged increase in catchability (e.g. UF-k2; Figures 7–8 and Figures S2–S5 in the Supplementary material).

When an incomplete fourth survey was added in the Atlantic sharpnose shark simulation, both CI and DFA assessment performance improved. This was likely due to the fourth survey having constant catchability, thereby providing more support of the true underlying abundance pattern. Further, to account for missing data, selectivity was estimated with two time-blocks, which likely increased the flexibility in the DFA assessment and improved DFA assessment performance.

Presence of conflicting survey indices

<u>Atlantic sharpnose shark</u>

In the case of conflicting survey indices, CI and DFA assessments generally maintained similar performance than in scenarios where catchability was constant for all survey indices (Figures 7–8). Notable exceptions were when underlying population abundance was constant (ConstF), and when the survey with the smallest *CV* underwent a shift in abundance that was in the opposite direction of the stock abundance trend (i.e. increasing catchability when stock size was decreasing; IncF_k2, IncF_g2, ConstF_k2, constF_g2). Across combinations of shifts in

catchability and survey-specific variability, CI assessment parameter estimates were generally more variable (longer violins), while DFA assessment parameter estimates were more consistent across scenarios (shorter violins; Figures 7–8 and Figures S2–S5 in the Supplementary material).

Sandbar shark

The sandbar shark simulation more thoroughly examined the effects of including conflicting survey indices within a stock assessment model by examining the number of surveys that were in conflict with the predominant trend (zero – three out of seven) and the directionality of the pattern of changing catchability (up, down). The DFA assessment generally performed more accurately across most scenarios than the CI assessment (Figures 9–10 and Figures S6–S9 in the Supplementary material). As in the Atlantic sharpnose shark simulation, DFA assessment estimates were more precise than CI assessment estimates across simulations in which survey catchability varied over time. However, the DFA assessment results for the sandbar shark were more variable under scenarios in which three surveys underwent shifts in catchability (namely C3_mix; Figures 9–10).

As the number of surveys that underwent shifts in catchability increased, CI and DFA assessment performance generally decreased, especially when three out of seven surveys were conflicting (Figures 7–8). Both the CI and DFA assessments performed more poorly in scenarios in which surveys experienced an increasing catchability pattern (Figures 9–10 and Figures S6–S9 in the Supplementary material). Note that an increasing shift in catchability is acting in the opposite direction of the predominant trend in the population (decreasing). Similarly, the DFA assessment experienced more variability across scenarios when three surveys experienced shifts in catchability and at least two of those surveys experienced increasing shifts in catchability (Figures 9–10 and Figures S6–S9 in the Supplementary material). This follows from Peterson *et al*. (2021), where DFA performance was poorer when shifts in catchability were in the opposite direction of the underlying population trend.

Missing data

The CI and DFA assessments fairly closely estimated depletion (Figure 9) and F_{final} (Figure 10) in all scenarios. Assessment performance with missing data was similar to assessment performance for complete data in both the CI and DFA assessments when zero, one, or two surveys underwent shifts in catchability. When three surveys were in conflict, assessment performance declined in both CI and DFA assessments when data were complete (Figures 9–10). Both CI and DFA assessments more accurately estimated parameters of interest in the missing data scenarios than did their complete data counterparts, indicating that the missing data may have allowed the model to rely less on relative abundance information and more on other pieces of information in those years. Length compositions were weighted differently in the CI and DFA assessments, and missing data DFA model runs also include time-blocks in selectivity, while the complete data DFA model runs do not, which may contribute to the differing results.

Interestingly, despite poorer DFA performance in missing data scenarios than in the complete data scenarios (Peterson *et al*., 2021), when three surveys were conflicting, DFA assessments with missing data more accurately estimated relative depletion and F_{final} than did the respective scenarios wherein data were complete (Figures 9–10 and Figures S6–S9 in the Supplementary material). Additionally, DFA assessments estimated depletion (and to a lesser extent F_{final}) better than CI assessments when zero, one, or two surveys underwent shifts in catchability for both complete and missing data (Figures 9–10). This is likely due to the

selectivity time-blocks implemented in the 'missing data' DFA assessment and larger uncertainty around the DFA trend, which allowed the assessment model to more heavily rely on other information within the model to produce a more realistic estimate of depletion and F_{final}.

Discussion

Main findings

Performance of CI and DFA assessments depended on underlying abundance scenarios, the direction and magnitude of changes in catchability combined with the survey variability, the number of surveys that conflicted, and whether the survey indices contained missing years of data. Comparable to findings by Wilberg and Bence (2006), failing to account for changes in survey catchability in the EM when they occurred in the OM resulted in biased parameter estimates in the CI and DFA assessments (i.e. interquartile range of relative parameter estimate did not include 1). In scenarios where survey catchability increased over time, relative depletion estimated from CI assessments was generally biased high, a result also found in Wilberg and Bence (2006). When the underlying population abundance was constant (Atlantic sharpnose shark ConstF), neither CI nor DFA assessments performed well when catchability was time-varying. Statistical models are known to struggle to fit data that do not exhibit sufficient contrast (Hilborn and Walters, 1992).

Despite these effects on DFA assessment performance, under scenarios in which DFA accurately predicted the underlying trend in the population (Peterson *et al*., 2021), DFA assessment performance was comparable to CI assessment performance overall. Dynamic factor analysis performs poorly when the underlying population is constant and multiple surveys undergo shifts in catchability (Peterson *et al*., 2021), and, consequently, the DFA assessment performance was also poor in these circumstances. Further, we note that by critically analyzing DFA results prior to including them in an assessment (e.g. whether the DFA trend is entirely inconsistent with remaining available data, including catches, length compositions, and life history strategy), many unreasonable and outlying assessment results may be avoided in practice. In this way, assessments may benefit from using DFA as a diagnostic technique when multiple conflicting indices are present, as DFA proved useful in elucidating underlying abundance trends from a collection of disagreeing survey indices (Peterson *et al*., 2021).

We applied DFA across surveys with various selectivities, an approach which has generally been discouraged (consider discussion within Conn, 2010a). However, the consistency between CI and DFA assessment results lends support for our approach to building the DFA assessment. Through DFA rescaling, we were able to backtransform the resulting DFA trend out from 'log-space' and preserve the lognormal error structure of the survey indices. We modeled selectivity by implementing a random-walk selectivity curve to approximate the average selectivity of each survey, including time-blocks to account for years with missing data, and we used factor loadings to weight mean length compositions within the DFA assessments. Any lingering concerns of combining indices with variable selectivity patterns could be alleviated by applying DFA to recruitment indices, for which length composition data are not required. Further, the DFA assessments may have an advantage over the CI approach by allowing the index selectivity to be estimated using a random walk. Modeling selectivity with a more flexible curve may have improved CI assessment results; however, there were insufficient length composition observations to model each fleet's selectivity using a random walk.

Ultimately, while there does not appear to be overwhelmingly clear evidence that using DFA to reconcile multiple survey indices of abundance prior to a stock assessment vastly

improves relevant parameter estimates in all simulated cases, DFA produced comparable or improved estimates in many scenarios and may serve additional purposes within a stock assessment. This approach of reconciling indices of abundance prior to an assessment model eliminates the potential of multiple, incongruous states of nature implied by making assumptions about the indices of abundance (as in the sandbar shark assessment in which all indices were included in the model as a base case, while increasing indices and decreasing indices were included as unique sensitivity runs, each suggesting a different stock status; SEDAR, 2017). Simplification of the input abundance indices to a single, stock-wide index may be considered a simpler assessment model which, when producing the same result as a more complex model, could be deemed preferable (Adkison, 2009).

Estimated parameters from DFA assessments were more consistently (precisely) estimated across variable simulation scenarios than those estimated from CI assessments. The DFA trend also has reduced interannual variability compared to the very large levels in corresponding individual indices, which are often incompatible with the slow life history of sharks in general and violate the assumption that the indices are proportional to population abundance (Cortés *et al*., 2015). Accordingly, DFA assessment results were largely buffered against biases introduced in estimated depletion and fishing mortality in the final year of fishing from time-varying catchability that were evident in the CI assessment results; though, this buffer did not hold when three surveys underwent time-varying catchability in the complete data sandbar shark simulation scenarios. For example, in many sandbar shark scenarios, the CI assessment predicted an increase in stock abundance at the end of the simulation, likely due to conflicting indices providing evidence that the stock was increasing in abundance. This predicted increase led to overestimated relative depletion by the CI assessments. These observations were slightly exacerbated in the complete data scenario, likely because the complete indices displayed a more complete and drastic contradiction to the underlying abundance when catchability was increasing or decreasing over time. The DFA reconciled these contrasting indices prior to the assessment, preventing discordant indices from impacting the DFA assessment to the same extent as in the CI assessments.

We also note that there were several scenarios in which DFA assessment estimates were more accurate than corresponding CI assessment estimates. Simply, when DFA 'works' for reconciling conflicting survey indices (assuming sufficient contrast in the stock abundance, where most survey indices are informative; see Peterson *et al*., 2021), using DFA results in an assessment generally 'works.' There are scenarios, however, in which DFA may be an inappropriate treatment of the data. Particularly, where data-limitation is not a concern, alternate spatio-temporal approaches (e.g. Thorson *et al*., 2015) may offer a more appropriate and informative treatment of the data. Further, when all surveys underwent shifts in catchability such that no index was representative of abundance, DFA was generally unable to recover the latent trend in stock abundance (Peterson *et al*., 2021). Ultimately, DFA results should be realistic and consistent with other information related to the stock (Peterson *et al*., 2021). Nevertheless, in the proper context, as guided by our simulation (appropriately applying DFA with our rescaling approach), while accounting for survey selectivity and length frequency observations, survey reconciliation approaches may serve as a valuable tool in fishery population dynamics analyses.

We also recognize the added potential value that DFA can serve within the stock assessment and management framework, apart from its use in generating a single index of abundance. DFA could be used as a valuable model simplification diagnostic tool in situations in which multiple conflicting indices induce confusion regarding the state of the system. Given the

demand for standardized diagnostics in integrated, age-structured stock assessment models, in addition to hierarchical cluster analysis (e.g. Courtney, 2017) or where hierarchical clustering breaks down, DFA could be useful in reducing the number of model runs required to evaluate model sensitivity to conflicting indices of abundance. Factor loadings (Γ_i) represent a way to separate input survey indices into agreeing and disagreeing groups (Zuur *et al*., 2007). An assessment may inherently benefit from a clearer understanding of the abundance pattern of the resource, *a priori*. Further, many index-based assessment and management procedure approaches are utilized in data-limited fisheries or to provide interim advice between full stock assessment years (Geromont and Butterworth, 2015; Huynh *et al.*, 2020). In these circumstances, it seems natural that use of a stock-wide DFA trend may be preferable to a geographically isolated index of abundance. We encourage future exploration of these concepts. Underlying population structure, missing data, and the interaction between conflicting surveys with decreased variability had the greatest effect on DFA assessment performance relative to CI assessment performance. Estimated quantities from DFA assessments appeared to be more precise than CI assessment results across variations in survey catchability. However, the precision of DFA assessment estimates tended to degrade as more surveys experienced trends in catchability (in the sandbar shark simulation). In scenarios where the underlying abundance provided good contrast (e.g. Atlantic sharpnose – UF and sandbar shark scenarios), DFA assessments generally produced better estimates of key stock status parameters. Assessment estimates were generally better in missing data scenarios (e.g. Atlantic sharpnose four surveys and sandbar shark missing data) in both CI and DFA assessments. DFA frequently outperformed CI assessments where missing data were present, and DFA assessments generally produced more accurate parameter estimates when data were complete. Nevertheless, there were scenarios in which CI assessments performed better (e.g. where surveys with the lowest CV underwent shifts in catchability in the opposite direction as the underlying trend in abundance; see SB29 and SB38 results in Figures S6–S9 in the Supplementary material).

General comments

A pillar of integrated analysis is the notion that data should be manipulated or treated as little as possible prior to inclusion into an assessment (Maunder and Punt, 2013; Methot and Wetzel, 2013). Utilization of raw data in integrated analysis aims to reduce loss of information, increase intuition and understandability with respect to diagnostics and likelihood functions, ensure logical consistencies, and most importantly, ensure uncertainty propagation and treatment (Maunder and Punt, 2013). Further, concerns regarding using an index reconciliation approach prior to an assessment model include correcting for variable survey catchabilities (Conn, 2010a; which also applies to individual indices), appropriate estimation of the selectivity curve (Azevedo *et al*., 2008; Conn, 2010a), and propagation of uncertainty (Maunder and Punt, 2013).

Then, why consider an approach that only increases pre-assessment data manipulation? We would argue that while it has generally been shown that manipulating data prior to input into stock assessments is not advisable, we need to be aware of the broader principles underpinning stock assessments and be mindful of the specific purpose that each data component is intended to achieve. At the core of an assessment, survey indices are intended to provide baseline trends in relative abundance, catches are used to provide the scale of abundance, life history information (i.e. reproductive rates, stock–recruit relationship, growth rates, natural mortality) provides the degree of vulnerability of the stock, and age, or more commonly, length frequencies, are used to

partition total abundance, as well as catch composition, into their representative size or age structure.

If we have spatially fragmented and highly uncertain survey indices that are no longer representative of the overall underlying trend in the resource, these indices are no longer serving their purpose within a stock assessment. By synthesizing these data prior to inclusion into an assessment, we are not carelessly undermining the integrated analysis framework, but rather, we are mindfully considering that, in certain instances, multiple conflicting indices may be an inappropriate representation of our stock. Data that are not representative of the resource should not be included in an assessment, but when given two opposing indices of abundance, it is generally not clear which (if either) index is a suitable representation of the underlying stock dynamics. Tools such as DFA can provide clarity in deciphering which survey indices may be relevant indicators of stock abundance via analysis of factor loadings.

In our analyses, we were able to address the issues identified above and partially alleviate some concerns regarding our additional pre-treatment of survey data. Through the process of rescaling survey indices (the relatively complex process of *Z*-scoring indices prior to fitting the DFA model), we corrected for the variable catchabilities across surveys. As catchability is simply a multiplicative constant, we removed the effect of multiplying constants of various magnitudes to each survey through standardization. This approach would not work for an index of abundance for which catchability is known to change over time; in those instances, attempts should be made to accommodate changes in catchability within survey index standardization approaches (e.g. Hinton and Maunder, 2003; Maunder and Punt, 2004) or within the stock assessment model (Wilberg *et al*., 2010), if possible. However, our results demonstrate that even when catchability is unknown and changing in a distinct pattern over time, DFA can provide realistic and relevant results.

We allowed the sandbar shark assessment model to internally estimate an age-based, random-walk selectivity function based on weighted length data, permitting the selectivity curve to follow unconventional forms. Length compositions from each survey were weighted by the relative magnitude of the corresponding factor loading. Effectively, this assumes that length observations from a survey that more strongly explained the resulting DFA trend are more informative, and surveys that did not contribute to the DFA trend were not used to estimate DFA trend selectivity. Hence, we accounted for the variable gear selectivities that are combined to produce the DFA-predicted trend.

As a state-space process, DFA is capable of estimating both observation and process errors, which represents a mechanism for uncertainty propagation. In our approach, we can specify known uncertainty attributed to each survey index into the DFA model as the known observation error and estimate a single trend that includes known observation error and estimated process error. Resulting standard errors of the DFA trend reflect this additional uncertainty. We can also allow the DFA model to internally estimate observation error and/or select from multiple different structural variance-covariance forms (Holmes *et al.*, 2014).

It is also worth noting that many integrated assessment approaches, including Stock Synthesis, require relative abundance information to be input into the model in an already modified form. Rather than raw cpue data, Stock Synthesis requires relative abundance information be input into the assessment model as an index, which requires manipulation of raw catch data (Maunder and Punt, 2013). We address many concerns related to data reconciliation prior to an assessment model and ultimately argue that supplying meaningful trends in relative abundance to an assessment is more important than keeping input data in raw form.

Generating a single trend in stock abundance also provides further clarity on the status of the resource, which will prove useful both within and outside the context of a stock assessment (Peterson *et al*., 2021). Disagreeing survey indices usually violate fishery-independent sampling assumptions, resulting in temporal trends that are not always representative of the abundance of the stock (Maunder *et al*., 2006; Wilberg *et al*., 2010; Maunder and Piner, 2017). By integrating spatiotemporally incomplete survey indices prior to an assessment, we obtain a stock-wide trend in abundance, which fulfills the goals of a proper survey and more consistently aligns with the purpose of a stock assessment. Proper interpretation of the trend in the resource prior to a stock assessment may also serve to improve stakeholder understanding and facilitate transparency of the assessment process. Thus, our prior treatment could be considered a more appropriate utilization of data.

Relevance

The disparate and conflicting nature of the indices of abundance available for coastal and other shark assessments is a well-recognized concern and hinders our understanding of the status of these stocks (Cortés, 2011; Cortés *et al*., 2015; Peterson *et al*., 2017). Research on integrating local abundance indices into a global index and identifying indices that contribute the most information to stock-wide trends has consistently been requested (ASMFC, 2013; SEDAR, 2013). Past coastal shark assessments have accordingly experimented with survey index reconciliation approaches (Conn, 2010b) in sensitivity runs (SEDAR, 2011).

The results of our research show that reconciling survey indices using DFA in the context of stock assessment is a justifiable exercise to improve understanding of the relative abundance of the stock. We present solutions to challenges that may be encountered in such analyses, including rescaling DFA results and treatment of selectivity for combined length or age compositions. Although we focused on coastal shark species in this simulation exercise, we found that this approach was applicable for two species of different life history and data availability. As such, we expect that the same protocol will be generalizable to other stocks.

We further highlight the value that DFA may have within the broader stock assessment framework. Consider index-based, data-limited assessment approaches, in which an index of abundance is a key input into an empirical assessment approach (e.g. Brooks *et al*., 2010; An Index Method, https://nmfs-fish-tools.github.io/). In these data-limited assessment methods, instead of inputting a single index that may be chosen somewhat arbitrarily, a DFA trend could be used instead (see Cortés and Brooks, 2018 for a similar application). These arguments extend to empirically based management procedures (Geromont *et al.*, 1999; Geromont and Butterworth, 2015) or interim assessment approaches (Huynh *et al*., 2020) used to update Annual Catch Limits (ACLs), wherein a stock-wide index compilation may produce more realistic or appropriate advice for an entire stock.

Supplementary material

The following Supplementary material including details of the DFA rescaling approach, treatment of length composition for the DFA estimation model, ancillary results, and additional tables and figures depicting full results for each simulation scenario is available at *ICESJMS* online.

Acknowledgements

We thank Andrew Scheld and Kyle Shertzer for reviews of earlier versions of this manuscript. CDP was funded by the NMFS Sea Grant Population and Ecosystem Dynamics Fellowship (NA17OAR4170242). This paper is Contribution No. 4045 of the Virginia Institute of Marine Science, William & Mary.

Data availability

There are no new data associated with this article. R code is available via https://github.com/cassidydpeterson/DFA_Simulation_and_Assessment.

References

Adkison, M. D. 2009. Drawbacks of complex models in frequentist and Bayesian approaches to natural-resource management. Ecological Applications, 19: 198–205.

ASMFC. 2013. Research Priorities and Recommendations to Support Interjurisdictional Fisheries Management: Coastal Sharks. ICES Document Special Report No. 89. 58 pp.

Azevedo, M., Duarte, R., Carador, F., Sousa, P., Fariña, C., Sampedro, P., Landa, J., *et al*. 2008. Application of dynamic factor analysis in the assessment of Iberian anglerfish stocks. ICES Journal of Marine Science, 65: 1365–1369.

Baremore, I. E., and Hale, L. F. 2012. Reproduction of the Sandbar Shark in the Western North Atlantic Ocean and Gulf of Mexico. Marine and Coastal Fisheries, 4: 560–572.

Bertsekas, D. P., and Tsitsiklis, J. N. 2002. Introduction to Probability, Athena Scientific, Belmont, MA, USA. 69 pp.

Bigelow, H. B., and Schroeder, W. C. 1948. Fishes of the Western North Atlantic, Sears Foundation for Marine Research, Yale University, New Haven.

Brewster-Geisz, K. K., and Miller, T. J. 2000. Management of the sandbar shark, *Carcharhinus plumbeus*: implications of a stage-based model. Fishery Bulletin, 98: 236–249.

Brooks, E. N., Powers, J. E., and Cortés, E. 2010. Analytical reference points for age-structured models: application to data-poor fisheries. ICES Journal of Marine Science, 67: 165–175.

Buckland, S. T., Studeny, A. C., Magurran, A. E., Illian, J. B., and Newson, S. E. 2011. The geometric mean of relative abundance indices: a biodiversity measure with a difference. Ecosphere, 2: 100.

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., *et al.* 2021. A cookbook for using model diagnostics in integrated stock assessments. Fisheries Research, 240: 105959.

Castro, J. I. 2009. Observations on the reproductive cycles of some viviparous North American sharks. aqua, International Journal of Ichthyology, 15: 205–222.

Conn, P. B. 2010a. Hierarchical analysis of multiple noisy abundance indices. Canadian Journal of Fisheries and Aquatic Sciences, 67: 108–120.

Conn, P. B. 2010b. Hierarchical analysis of blacknose, sandbar, and dusky shark CPUE indices. Southeast Data Assessment and Review 21 Assessment Workshop for blacknose, sandbar, and dusky shark, SEDAR 21-AP-01. 20 pp.

Cook, R. 2010. Reviewers Report of SEDAR 21 Data Workshop (DW) HMS sandbar, dusky, and blacknose shark assessment. 24 pp.

Cope, J. M. 2013. Implementing a statistical catch-at-age model (Stock Synthesis) as a tool for deriving overfishing limits in data-limited situations. Fisheries Research, 142: 3–14.

Cortés, E. 2011. An overview of approaches used to assess the status of shark populations: experiences from the USA and ICCAT in the Atlantic Ocean. Document IOTC-2011-WPEB07-25.

Cortés, E., and Brooks, E. N. 2018. Stock status and reference points for sharks using data-limited methods and life history. Fish and Fisheries, 19: 1110–1129.

Cortés, E., Brooks, E. N., and Shertzer, K. W. 2015. Risk assessment of cartilaginous fish populations. ICES Journal of Marine Science, 72: 1057–1068.

Courtney, D. 2016. Preliminary stock synthesis (SS3) model runs conducted for North Atlantic blue shark. Collective Volume of Scientific Papers ICCAT, 72: 1186–1232.

Courtney, D. 2017. Example implementation of a hierarchical cluster analysis and cross-correlations of selected CPUE indices for the SEDAR 54 assessment. SEDAR54- WP-06. SEDAR, North Charleston, SC. 11 pp.

Courtney, D., Cortés, E., and Zhang, X. 2017b. Stock synthesis (SS3) model runs conducted for North Atlantic shortfin mako shark. Collective Volume of Scientific Papers ICCAT, 74: 1759–1821.

Courtney, D., Cortés, E., Zhang, X., and Carvalho, F. 2017a. Stock synthesis model sensitivity to data weighting: an example from preliminary model runs previously conducted for North Atlantic blue shark. Collective Volume of Scientific Papers ICCAT, 73: 2860–2890.

Courtney, D. L., Adkison, M. D., and Sigler, M. F. 2016. Risk analysis of plausible incidental exploitation rates for the Pacific Sleeper Shark, a data-poor species in the Gulf of Alaska. North American Journal of Fisheries Management, 36: 523–548.

Davis, M. M., Suárez-Moo, P. D. J., and Daly-Engel, T. S. 2019. Genetic structure and congeneric range overlap among sharpnose sharks (genus *Rhizoprionodon*) in the Northwest Atlantic Ocean. Canadian Journal of Fisheries and Aquatic Sciences, 76: 1203–1211.

Field, I. C., Meekan, M. G., Buckworth, R. C., and Bradshaw, C. J. A. 2009. Susceptibility of sharks, rays and chimaeras to global extinction. Advances in Marine Biology, 56: 275–363.

Francis, R. I. C. C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 68: 1124–1138.

Francis, R. I. C. C. 2017. Revisiting data weighting in fisheries stock assessment models. Fisheries Research, 192: 5–15.

Frazier, B. S., Driggers Iii, W. B., and Ulrich, G. F. 2014. Longevity of Atlantic Sharpnose Sharks *Rhizoprionodon terraenovae* and Blacknose Sharks *Carcharhinus acronotus* in the western North Atlantic Ocean based on tag-recapture data and direct age estimates. F1000Research, 3: 190.

Geromont, H. F., and Butterworth, D. S. 2015. Generic management procedures for data-poor fisheries: forecasting with few data. ICES Journal of Marine Science, 72: 251–261.

Geromont, H. F., de Oliveira, J. A. A., Johnston, S. J., and Cunningham, C. L. 1999. Development and application of management procedures for fisheries in southern Africa. ICES Journal of Marine Science, 56: 952–966.

Heist, E. J., Graves, J. E., and Musick, J. A. 1995. Population genetics of the sandbar shark (*Carcharhinus plumbeus*) in the Gulf of Mexico and Mid-Atlantic Bight. Copeia, 1995: 555–562.

Hilborn, R., and Walters, C. J. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty, Chapman & Hall, London.

Hinton, M. G., and Maunder, M. N. 2003. Methods for standardizing CPUE and how to select among them. Document SCTB16 Working Paper MWG-7. Inter-American Tropical Tuna Commission, La Jolla, CA. 11 pp.

Holmes, E. E., Ward, E. J., and Scheuerell, M. D. 2014. Analysis of multivariate time-series using the MARSS package. NOAA Fisheries, Northwest Fisheries Science Center, Seattle, WA. 333 pp.

Huynh, Q. C., Hordyk, A. R., Forrest, R. E., Porch, C. E., Anderson, S. C., and Carruthers, T. R. 2020. The interim management procedure approach for assessed stocks: Responsive management advice and lower assessment frequency. Fish and Fisheries, 21: 663–679.

Kohler, N. E., Casey, J. G., and Turner, P. A. 1998. NMFS Cooperative Shark Tagging Program, 1962-93: An atlas of shark tagging and recapture data. Marine Fisheries Review, 60: 1–87.

Loefer, J. K., and Sedberry, G. R. 2003. Life history of the Atlantic sharpnose shark (*Rhizoprionodon terraenovae*) (Richardson, 1836) off the southeastern United States. Fishery Bulletin, 101: 75–88.

Maunder, M. N. 2001. A general framework for integrating the standardization of catch per unit of effort into stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 58: 795–803.

Maunder, M. N., and Piner, K. R. 2017. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Fisheries Research, 192: 16–27.

Maunder, M. N., and Punt, A. E. 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research, 70: 141–159.

Maunder, M. N., and Punt, A. E. 2013. A review of integrated analysis in fisheries stock assessment. Fisheries Research, 142: 61–74.

Maunder, M. N., Sibert, J. R., Fonteneau, A., Hampton, J., Kleiber, P., and Harley, S. J. 2006. Interpreting catch per unit effort data to assess the status of individual stocks and communities. ICES Journal of Marine Science, 63: 1373–1385.

Methot, R. D. 2013. User Manual for Stock Synthesis Version 3.24s. NOAA Fisheries Seattle, WA. 152 pp.

Methot, R. D., and Taylor, I. G. 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 68: 1744–1760.

Methot, R. D., and Wetzel, C. R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research, 142: 86–99.

Ono, K., Ianelli, J. N., McGilliard, C. R., and Punt, A. E. 2018. Integrating data from multiple surveys and accounting for spatio-temporal correlation to index the abundance of juvenile Pacific halibut in Alaska. ICES Journal of Marine Science, 75: 572–584.

Peterson, C. D., Belcher, C. N., Bethea, D. M., Driggers, W. B., Frazier, B. S., and Latour, R. J. 2017. Preliminary recovery of coastal sharks in the south-east United States. Fish and Fisheries, 18: 845–859.

Peterson, C. D., Wilberg, M. J., Cortés, E., and Latour, R. J. 2021. Dynamic factor analysis to reconcile conflicting survey indices of abundance. ICES Journal of Marine Science: 1–19.

R Core Team. 2019. R: A language and environment for statistical computing. version 3.6.2 edn. R Foundation for Statistical Computing, Vienna, Austria.
Schnute, J. T., and Hilborn, R. 1993. Analysis of contradictory data sources in fish stock assessment. Canadian Journal of Fisheries and Aquatic Sciences, 50: 1916–1923.
SEDAR. 2011. SEDAR 21 HMS Sandbar Shark Stock Assessment Report.
SEDAR. 2013. SEDAR 34 HMS Atlantic Sharpnose Shark Stock Assessment Report.
SEDAR. 2017. SEDAR 54 HMS Sandbar Shark Stock Assessment Report.
Sminkey, T. R., and Musick, J. A. 1995. Age and growth of the sandbar shark, *Carcharhinus plumbeus*, before and after population depletion. Copeia, 1995: 871–883.
Springer, S. 1960. Natural history of the sandbar shark *Eulamia milberti*. Fishery Bulletin, 61: 1–38.
Stevens, J. D., Bonfil, R., Dulvy, N. K., and Walker, P. A. 2000. The effects of fishing on sharks, rays, and chimaeras (chondrichthyans), and the implications for marine ecosystems. ICES Journal of Marine Science, 57: 476–494.
Taylor, I. G., Doering, K. L., Johnson, K. F., Wetzel, C. R., and Stewart, I. J. 2021. Beyond visualizing catch-at-age models: Lessons learned from the r4ss package about software to support stock assessments. Fisheries Research, 239: 105924.
Taylor, I. G., Gertseva, V., Methot, R. D., and Maunder, M. N. 2013. A stock–recruitment relationship based on pre-recruit survival, illustrated with application to spiny dogfish shark. Fisheries Research, 142: 15–21.
Taylor, I. G., Stewart, I. J., Hicks, A. C., Garrison, T. M., Punt, A. E., Wallace, J. R., Wetzel, C. R., *et al*. 2019. r4ss: R Code for Stock Synthesis. R package version 1.36.1 edn.
Thorson, J. T., Shelton, A. O., Ward, E. J., and Skaug, H. J. 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES Journal of Marine Science, 72: 1297-1310.
Wetzel, C. R., and Punt, A. E. 2011. Performance of a fisheries catch-at-age model (Stock Synthesis) in data-limited situations. Marine and Freshwater Research, 62: 927–936.
Wilberg, M. J., and Bence, J. R. 2006. Performance of time-varying catchability estimators in statistical catch-at-age analysis. Canadian Journal of Fisheries and Aquatic Sciences, 63: 2275–2285.
Wilberg, M. J., Thorson, J. T., Linton, B. C., and Berkson, J. 2010. Incorporating time-varying catchability into population dynamic stock assessment models. Reviews in Fisheries Science, 18: 7–24.
Zuur, A. F., Fryer, R. J., Jolliffe, I. T., Dekker, R., and Beukema, J. J. 2003b. Estimating common trends in multivariate time series using dynamic factor analysis. Environmetrics, 14: 665–685.
Zuur, A. F., Ieno, E. N., and Smith, G. M. 2007. Analysing Ecological Data*,* Springer, New York, NY.
Zuur, A. F., Tuck, I. D., and Bailey, N. 2003a. Dynamic factor analysis to estimate common trends in fisheries time series. Canadian Journal of Fisheries and Aquatic Sciences, 60: 542–552.

Table 1. List of equations and parameter definitions.

1	$N_{s,a,y+1} = \begin{cases} R_{s,y+1} & a = \alpha \\ N_{s,a-1,y}\mathrm{e}^{-Z_{s,a,y}} & \alpha < a < A \\ \left(N_{s,A-1,y}\mathrm{e}^{-Z_{s,A-1,y}}\right) + \left(N_{s,A,y}\mathrm{e}^{-Z_{s,A,y}}\right) & a = A \end{cases}$	Age-structured numbers at age
2	$R_{s,y+1} = Npups_y \times e^{-M_{0,y}}$ $M_{0,y} \sim N\left\{\left[1 - \left(\frac{Npups_y}{Npups_0}\right)^{\beta} \times (Z_{min} - Z_0) + Z_0\right], 0.1\right\}$ $Npups_y = \sum_a N_{s=female,a,y} \times \frac{1}{RP} \times p_{f,a} \times f_a$ $p_{s=female,a} \sim N\left(p_{avg}{}_{s=female,a}, p_{avg}{}_{s=female,a} \times 0.01\right)$ $f_a \sim N\left(f_{avg}{}_a, f_{avg}{}_a \times 0.1\right)$	LFSR recruitment
3	$C_{s,a,y,i} = N_{s,a,y} \times \frac{F_{s,a,y,i}}{Z_{s,a,y}} \times (1 - e^{-Z_{s,a,y}})$ $F_{s,a,y,i} = sel_{s,a,i} \times F_{s,a,y} \times \delta_i$	Catch
4	$Z_{s,a,y} = \sum_{\forall i} F_{s,a,y,i} + M_{s,a}$ $M_{s,a} \sim N\left(M_{avg}{}_{s,a}, M_{avg}{}_{s,a} * 0.01\right)$	Total instantaneous mortality
5	$I_{s,a,y,j} = q_{y,j} v_{s,a,j} N_{s,a,y} \times \exp\left(\epsilon_{s,a,y,j} - \frac{\sigma^2_{s,a,y,j}}{2}\right)$ $\epsilon_{s,a,y,j} \sim N\left(0, \sigma_{s,a,y,j}\right)$ $\sigma_{s,a,y,j} = q_{y,j} v_{s,a,j} N_{s,a,y} CV_j$ $CV_j \sim U\left(CV_{avg_j} - 0.1, CV_{avg_j} + 0.1\right)$ $I_{y,j} = \sum_{\forall a} I_{s=female,a,y,j} + I_{s=male,a,y,j}$	Indices of abundance
6	$L_n = L_\infty\left(1 - e^{-K(jitter(a_n) - t_0)}\right)$ $L_{\infty_n} \sim N\left(L_{\infty_{avg}}, \sigma^2_{L_\infty}\right), \quad \sigma^2_{L_\infty} = 0.1 L_{\infty_{avg}}$ $K \sim N\left[K_{avg} + \left(\frac{\sigma_K^2}{\sigma^2_{L_\infty}}\right)\rho\left(L_{\infty_n} - L_{\infty_{avg}}\right), \sqrt{(1-\rho^2)(\sigma_K^2)^2}\right], \quad \sigma_K^2 = 0.1 K_{avg}$	von Bertalanffy length calculation
7	$y_t = \boldsymbol{\Gamma}\alpha_t + \varepsilon_t$, where $\varepsilon_t \sim MVN(0, \boldsymbol{H})$ $\alpha_t = \alpha_{t-1} + \eta_t$, where $\eta_t \sim MVN(0, \boldsymbol{Q})$	DFA model
Parameter definitions: abundance in numbers (N), recruits in numbers (R), total instantaneous mortality (Z), age of plus group (A), pre-recruits in numbers ($Npups$), natural mortality (M), reproductive periodicity (RP), proportion mature (p), fecundity (f), fishing mortality (F), catch in numbers (C), proportion of fishing mortality attributed to fleet (δ), index of abundance (I), catchability (q), vulnerability (v), length (L), von Bertalanffy growth parameters (L_∞ and K), correlation (ρ), rescaled index of abundance (y), factor loadings matrix ($\boldsymbol{\Gamma}$), DFA common trend (α) Subscripts: sex (s), age (a), age of plus group (A), year (y), fleet (i), survey (j), individual (n), time (t)		

Table 2. List of trials simulated for the Atlantic sharpnose shark. F is instantaneous fishing mortality pattern, I represents index of abundance indexed by survey number, CV is coefficient of variation, and q is catchability coefficient. Increasing and decreasing are indicated by Inc/Dec or ↑/↓. Trials were labeled according to the F pattern (constant – ConstF, increasing – IncF, decreasing – DecF, or increasing then decreasing – U F), changing catchability (constant – c, knife-edged – k, or gradual – g), and indexed by the survey-specific pattern (1, 2, or 3). Trials with 4 simulated surveys were indexed by 4.

Figure 4 plotting labels	**Trial name (3 surveys)**	**F**[a]	**I1 CV**	**I2 CV**	**I3 CV**	**I1 *q***[b]	**I2 *q***	**I3 *q***[b]	**I4 CV**	**I4 *q***	**Trial name (4 surveys)**[c]	**Figure 4 plotting labels**
Const F no change	ConstF_c1	Const F (F=0.2)	0.5	0.5	0.5	const *q*	const *q*	const *q*	0.5	const *q*	ConstF_c1_4	Const F no change
	ConstF_c2	Const F (F=0.2)	0.3	0.5	0.7	const *q*	const *q*	const *q*	0.5	const *q*	ConstF_c2_4	
	ConstF_c3	Const F (F=0.2)	0.7	0.5	0.3	const *q*	const *q*	const *q*	0.5	const *q*	ConstF_c3_4	
Const F knife-edge	ConstF_k1	Const F (F=0.2)	0.5	0.5	0.5	knife ↑*q*	const *q*	knife ↓*q*	0.5	const *q*	ConstF_k1_4	Const F knife-edge
	ConstF_k2	Const F (F=0.2)	0.3	0.5	0.7	knife ↑*q*	const *q*	knife ↓*q*	0.5	const *q*	ConstF_k2_4	
	ConstF_k3	Const F (F=0.2)	0.7	0.5	0.3	knife ↑*q*	const *q*	knife ↓*q*	0.5	const *q*	ConstF_k3_4	
Const F gradual	ConstF_g1	Const F (F=0.2)	0.5	0.5	0.5	grad ↑*q*	const *q*	grad ↓*q*	0.5	const *q*	ConstF_g1_4	Const F gradual
	ConstF_g2	Const F (F=0.2)	0.3	0.5	0.7	grad ↑*q*	const *q*	grad ↓*q*	0.5	const *q*	ConstF_g2_4	
	ConstF_g3	Const F (F=0.2)	0.7	0.5	0.3	grad ↑*q*	const *q*	grad ↓*q*	0.5	const *q*	ConstF_g3_4	
Inc F no change	IncF_c1	↑F (F=0 / 0.4)	0.5	0.5	0.5	const *q*	const *q*	const *q*	0.5	const *q*	IncF_c1_4	Inc F no change
	IncF_c2	↑F (F=0 / 0.4)	0.3	0.5	0.7	const *q*	const *q*	const *q*	0.5	const *q*	IncF_c2_4	
	IncF_c3	↑F (F=0 / 0.4)	0.7	0.5	0.3	const *q*	const *q*	const *q*	0.5	const *q*	IncF_c3_4	
Inc F knife-edge	IncF_k1	↑F (F=0 / 0.4)	0.5	0.5	0.5	knife ↑*q*	const *q*	knife ↓*q*	0.5	const *q*	IncF_k1_4	Inc F knife-edge
	IncF_k2	↑F (F=0 / 0.4)	0.3	0.5	0.7	knife ↑*q*	const *q*	knife ↓*q*	0.5	const *q*	IncF_k2_4	
	IncF_k3	↑F (F=0 / 0.4)	0.7	0.5	0.3	knife ↑*q*	const *q*	knife ↓*q*	0.5	const *q*	IncF_k3_4	
Inc F gradual	IncF_g1	↑F (F=0 / 0.4)	0.5	0.5	0.5	grad ↑*q*	const *q*	grad ↓*q*	0.5	const *q*	IncF_g1_4	Inc F gradual
	IncF_g2	↑F (F=0 / 0.4)	0.3	0.5	0.7	grad ↑*q*	const *q*	grad ↓*q*	0.5	const *q*	IncF_g2_4	
	IncF_g3	↑F (F=0 / 0.4)	0.7	0.5	0.3	grad ↑*q*	const *q*	grad ↓*q*	0.5	const *q*	IncF_g3_4	
Dec F no change	DecF_c1	↓F (F=0.4 / 0)	0.5	0.5	0.5	const *q*	const *q*	const *q*	0.5	const *q*	DecF_c1_4	Dec F no change
	DecF_c2	↓F (F=0.4 / 0)	0.3	0.5	0.7	const *q*	const *q*	const *q*	0.5	const *q*	DecF_c2_4	
	DecF_c3	↓F (F=0.4 / 0)	0.7	0.5	0.3	const *q*	const *q*	const *q*	0.5	const *q*	DecF_c3_4	
Dec F knife-edge	DecF_k1	↓F (F=0.4 / 0)	0.5	0.5	0.5	knife ↑*q*	const *q*	knife ↓*q*	0.5	const *q*	DecF_k1_4	Dec F knife-edge
	DecF_k2	↓F (F=0.4 / 0)	0.3	0.5	0.7	knife ↑*q*	const *q*	knife ↓*q*	0.5	const *q*	DecF_k2_4	
	DecF_k3	↓F (F=0.4 / 0)	0.7	0.5	0.3	knife ↑*q*	const *q*	knife ↓*q*	0.5	const *q*	DecF_k3_4	
Dec F gradual	DecF_g1	↓F (F=0.4 / 0)	0.5	0.5	0.5	grad ↑*q*	const *q*	grad ↓*q*	0.5	const *q*	DecF_g1_4	Dec Fgradual
	DecF_g2	↓F (F=0.4 / 0)	0.3	0.5	0.7	grad ↑*q*	const *q*	grad ↓*q*	0.5	const *q*	DecF_g2_4	
	DecF_g3	↓F (F=0.4 / 0)	0.7	0.5	0.3	grad ↑*q*	const *q*	grad ↓*q*	0.5	const *q*	DecF_g3_4	
U F no change	UF_c1	UF (F=0 / 0.4 / 0.2 / 0.05)	0.5	0.5	0.5	const *q*	const *q*	const *q*	0.5	const *q*	UF_c1_4	U F no change
	UF_c2	UF (F=0 / 0.4 / 0.2 / 0.05)	0.3	0.5	0.7	const *q*	const *q*	const *q*	0.5	const *q*	UF_c2_4	

	UF_c3	UF (F=0 / 0.4 / 0.2 / 0.05)	0.7	0.5	0.3	const *q*	const *q*	const *q*	0.5	const *q*	UF_c3_4	
U F knife-edge	UF_k1	UF (F=0 / 0.4 / 0.2 / 0.05)	0.5	0.5	0.5	knife ↑*q*	const *q*	knife ↓*q*	0.5	const *q*	UF_k1_4	U F knife-edge
	UF_k2	UF (F=0 / 0.4 / 0.2 / 0.05)	0.3	0.5	0.7	knife ↑*q*	const *q*	knife ↓*q*	0.5	const *q*	UF_k2_4	
	UF_k3	UF (F=0 / 0.4 / 0.2 / 0.05)	0.7	0.5	0.3	knife ↑*q*	const *q*	knife ↓*q*	0.5	const *q*	UF_k3_4	
U F gradual	UF_g1	UF (F=0 / 0.4 / 0.2 / 0.05)	0.5	0.5	0.5	grad ↑*q*	const *q*	grad ↓*q*	0.5	const *q*	UF_g1_4	U F gradual
	UF_g2	UF (F=0 / 0.4 / 0.2 / 0.05)	0.3	0.5	0.7	grad ↑*q*	const *q*	grad ↓*q*	0.5	const *q*	UF_g2_4	
	UF_g3	UF (F=0 / 0.4 / 0.2 / 0.05)	0.7	0.5	0.3	grad ↑*q*	const *q*	grad ↓*q*	0.5	const *q*	UF_g3_4	
Simulation Notes:												
[a] shifts in *F* for ↑F and ↓F occurred in year 51, and shifts in UF occurred at years 41, 51, and 56												
[b] all knife-edged shifts in *q* occurred at year 51; gradual shifts in q_1 and q_3 spanned 15 and 10 years, respectively, starting at year 51												
[c] when present, 4th index started at year 55												

Table 3. List of trials simulated for the sandbar shark. 'Missing data?' indicates whether survey indices were complete or whether missing values were included to more accurately represent available information for the sandbar shark, *q* indicates catchability coefficient, and only one instantaneous fishing mortality (*F*) scenario was explored. ↑ represents increasing patterns in *q*, while ↓ represents decreasing patterns in *q*. Prefixes M and C indicate missing data and complete data scenarios, respectively.

Figure plotting labels	**Trial**	**Missing data?**	**F[a]**	***q* pattern**	**Trial**	**Missing data?**	**F[a]**	***q* pattern**	**Figure plotting labels**
M0	**SB1**	Yes	SB_F	const	**SB101**	No	SB_F	const	C0
M1_up	**SB2**	Yes	SB_F	↑*q* S1	**SB102**	No	SB_F	↑*q* S1	C1_up
	SB3	Yes	SB_F	↑*q* S2	**SB103**	No	SB_F	↑*q* S2	
	SB4	Yes	SB_F	↑*q* S3	**SB104**	No	SB_F	↑*q* S3	
	SB5	Yes	SB_F	↑*q* S4	**SB105**	No	SB_F	↑*q* S4	
	SB6	Yes	SB_F	↑*q* S5	**SB106**	No	SB_F	↑*q* S5	
	SB7	Yes	SB_F	↑*q* S6	**SB107**	No	SB_F	↑*q* S6	
	SB8	Yes	SB_F	↑*q* S7	**SB108**	No	SB_F	↑*q* S7	
M1_down	**SB9**	Yes	SB_F	↓*q* S1	**SB109**	No	SB_F	↓*q* S1	C1_down
	SB10	Yes	SB_F	↓*q* S2	**SB110**	No	SB_F	↓*q* S2	
	SB11	Yes	SB_F	↓*q* S3	**SB111**	No	SB_F	↓*q* S3	
	SB12	Yes	SB_F	↓*q* S4	**SB112**	No	SB_F	↓*q* S4	
	SB13	Yes	SB_F	↓*q* S5	**SB113**	No	SB_F	↓*q* S5	
	SB14	Yes	SB_F	↓*q* S6	**SB114**	No	SB_F	↓*q* S6	
	SB15	Yes	SB_F	↓*q* S7	**SB115**	No	SB_F	↓*q* S7	
M2_up	**SB16**	Yes	SB_F	↑*q* S1-S3	**SB116**	No	SB_F	↑*q* S1-S3	C2_up
	SB17	Yes	SB_F	↑*q* S3-S5	**SB117**	No	SB_F	↑*q* S3-S5	
	SB18	Yes	SB_F	↑*q* S5-S7	**SB118**	No	SB_F	↑*q* S5-S7	
M2_down	**SB19**	Yes	SB_F	↓*q* S1-S3	**SB119**	No	SB_F	↓*q* S1-S3	C2_down

	SB20	Yes	SB_F	↓q S3-S5	**SB120**	No	SB_F	↓q S3-S5	
	SB21	Yes	SB_F	↓q S5-S7	**SB121**	No	SB_F	↓q S5-S7	
M2_mix	**SB22**	Yes	SB_F	↑q S1,↓q S3	**SB122**	No	SB_F	↑q S1,↓q S3	C2_mix
	SB23	Yes	SB_F	↑q S3,↓q S5	**SB123**	No	SB_F	↑q S3,↓q S5	
	SB24	Yes	SB_F	↑q S5,↓q S7	**SB124**	No	SB_F	↑q S5,↓q S7	
M2_mix	**SB25**	Yes	SB_F	↓q S1,↑q S3	**SB125**	No	SB_F	↓q S1,↑q S3	C2_mix
	SB26	Yes	SB_F	↓q S3,↑q S5	**SB126**	No	SB_F	↓q S3,↑q S5	
	SB27	Yes	SB_F	↓q S5,↑q S7	**SB127**	No	SB_F	↓q S5,↑q S7	
M3_up	**SB28**	Yes	SB_F	↑q S1-S3-S5	**SB128**	No	SB_F	↑q S1-S3-S5	C3_up
	SB29	Yes	SB_F	↑q S2-S4-S6	**SB129**	No	SB_F	↑q S2-S4-S6	
	SB30	Yes	SB_F	↑q S3-S5-S7	**SB130**	No	SB_F	↑q S3-S5-S7	
M3_down	**SB31**	Yes	SB_F	↓q S1-S3-S5	**SB131**	No	SB_F	↓q S1-S3-S5	C3_down
	SB32	Yes	SB_F	↓q S2-S4-S6	**SB132**	No	SB_F	↓q S2-S4-S6	
	SB33	Yes	SB_F	↓q S3-S5-S7	**SB133**	No	SB_F	↓q S3-S5-S7	
M3_mix	**SB34**	Yes	SB_F	↑q S1, ↓q S3-S5	**SB134**	No	SB_F	↑q S1, ↓q S3-S5	C3_mix
	SB35	Yes	SB_F	↑q S2, ↓q S4-S6	**SB135**	No	SB_F	↑q S2, ↓q S4-S6	
	SB36	Yes	SB_F	↑q S3, ↓q S5-S7	**SB136**	No	SB_F	↑q S3, ↓q S5-S7	
M3_mix	**SB37**	Yes	SB_F	↓q S1, ↑q S3-S5	**SB137**	No	SB_F	↓q S1, ↑q S3-S5	C3_mix
	SB38	Yes	SB_F	↓q S2, ↑q S4-S6	**SB138**	No	SB_F	↓q S2, ↑q S4-S6	
	SB39	Yes	SB_F	↓q S3, ↑q S5-S7	**SB139**	No	SB_F	↓q S3, ↑q S5-S7	
Simulation Notes									
[a] SB F = 0 in years 1-45, 0.1 in years 46-55, 0.3 in years 56-65, 0.2 in years 66-75, 0.05 in years 76-100									
CV_1=0.38, CV_2=0.48, CV_3=0.65, CV_4=0.24, CV_5=0.30, CV_6=0.36, CV_7=0.40									

Figures

Figure 1. Example DFA model run for the first iteration of sandbar shark Trial SB129, including (a) input data as shown as mean-standardized survey indices, (b) corresponding factor loadings, denoting the strength of influence of the resulting DFA-predicted trend on each survey, (c) resulting DFA-predicted trend with 95% confidence intervals (CIs) in log space and the "true" simulated abundance trend log-transformed and rescaled superimposed, and (d) the backtransformed DFA trend with 95% CIs and the rescaled simulated abundance superimposed. Note that out of seven input survey indices, four do not undergo shift in catchability, while three experience increases in catchability (denoted by solid lines).

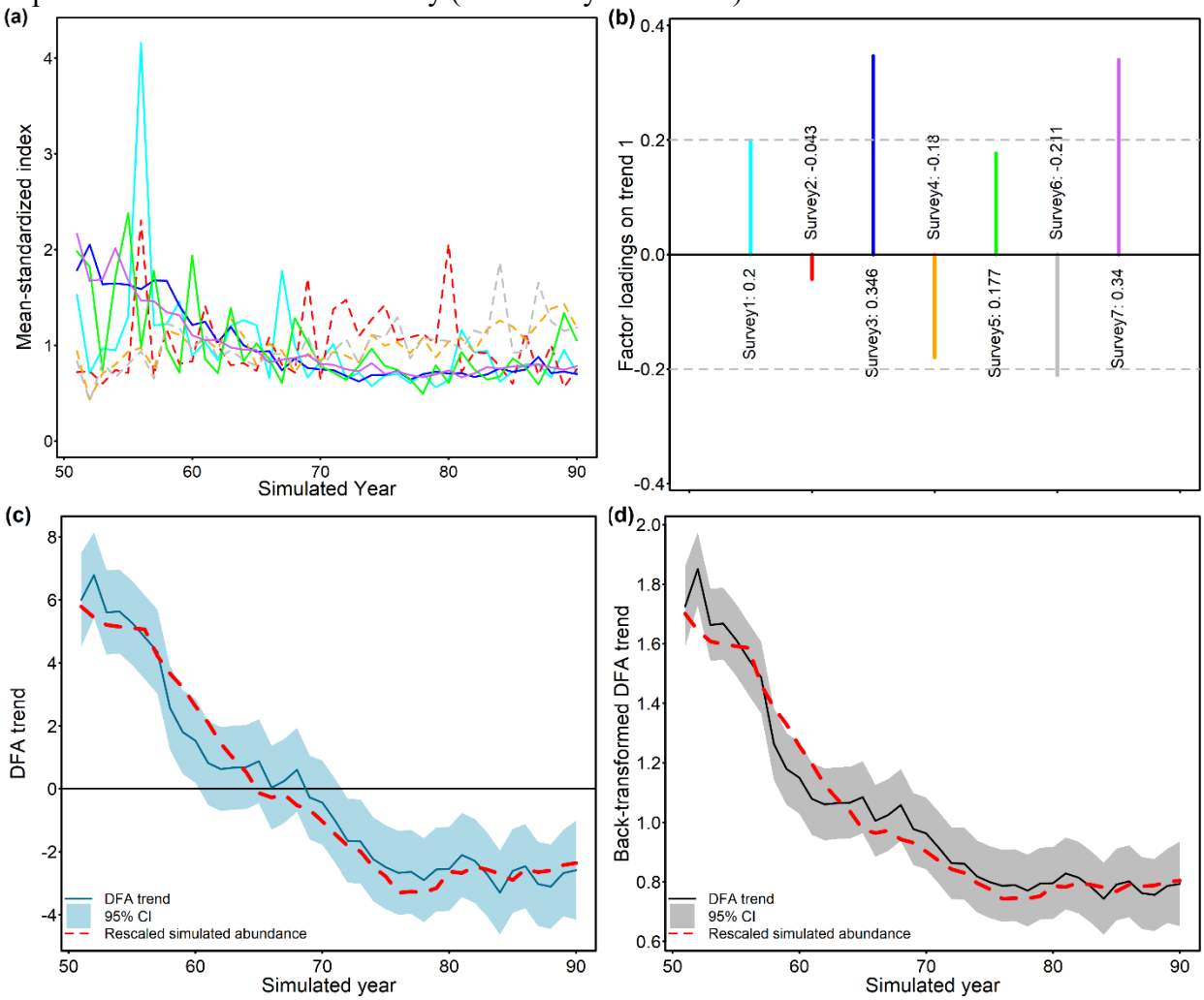


Figure 2. Selectivity curves of fishing fleets and surveys simulated in the Atlantic sharpnose (top left) and sandbar shark (top right and bottom) operating models (solid lines). Selectivity curves estimated within the estimating model are superimposed with dashed lines.

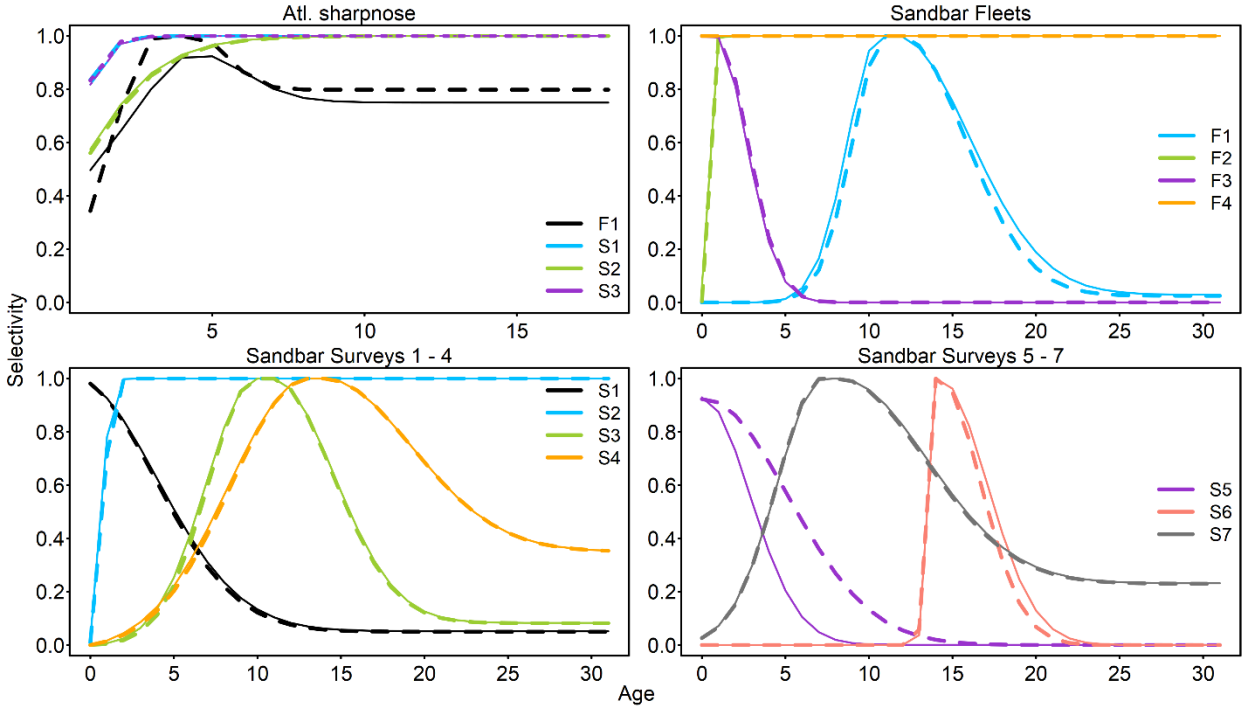


Figure 3. Taken from Peterson *et al.* (2021). Alternate simulation scenarios for the Atlantic sharpnose shark (left) and sandbar shark (right) including various time-varying catchability configurations (top row), fishing mortality patterns for Atlantic sharpnose shark simulations (bottom left), and available years of survey data in the 'Missing data' scenario for the sandbar shark simulation for each survey (bottom right).

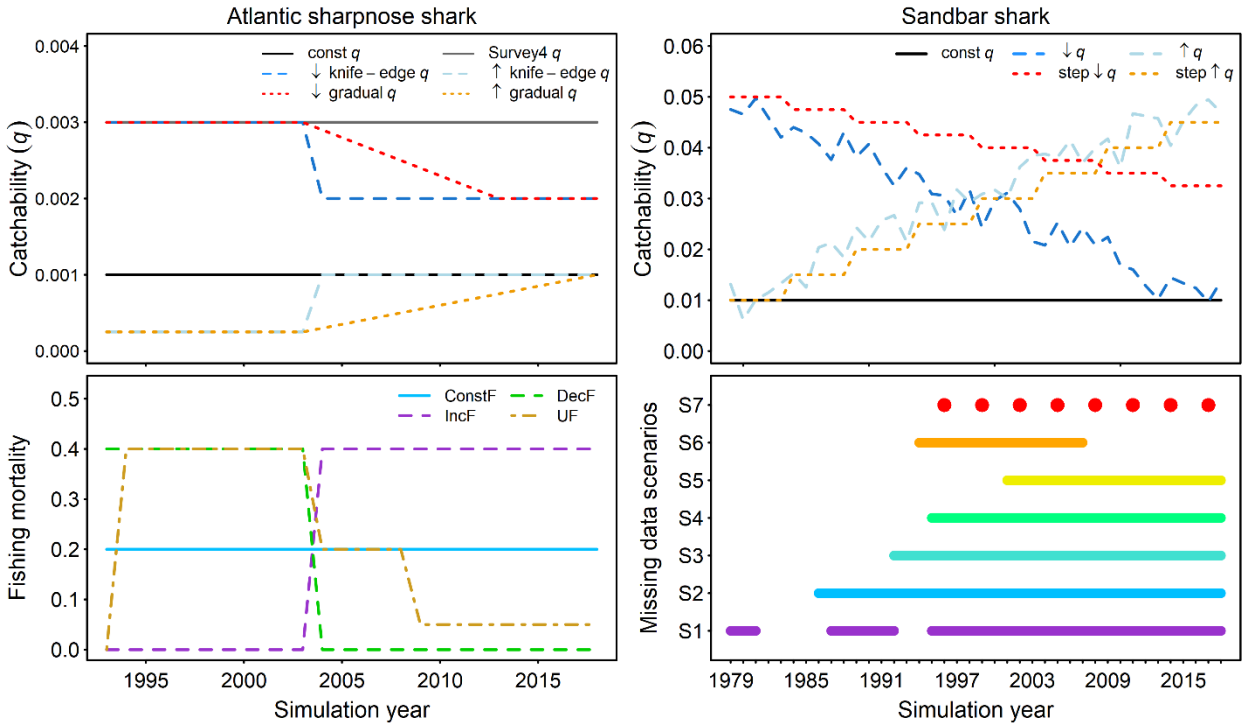


Figure 4. Survey-specific female selectivity as fitted from the sandbar shark CI assessment for iteration 1 of Trial SB101 (left). The average selectivity, calculated by averaging selectivity curves across surveys and scaling to a maximum selectivity of one, is superimposed. The random-walk both-sex selectivity as fitted within the sandbar shark DFA assessment for iteration 1 of Trial SB101 is shown on the right.

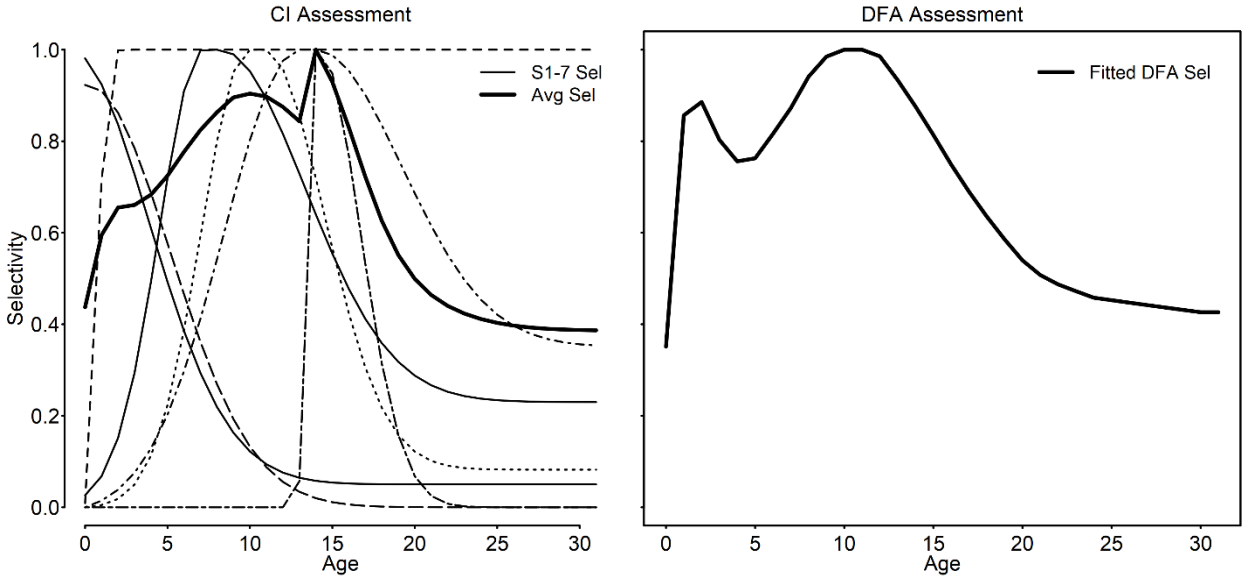


Figure 5. Example of the fleet-specific, time-aggregated length compositions from the CI assessment of the first iteration of the SB29 simulation scenario for the sandbar shark as plotted using r4ss (Taylor *et al*., 2021).

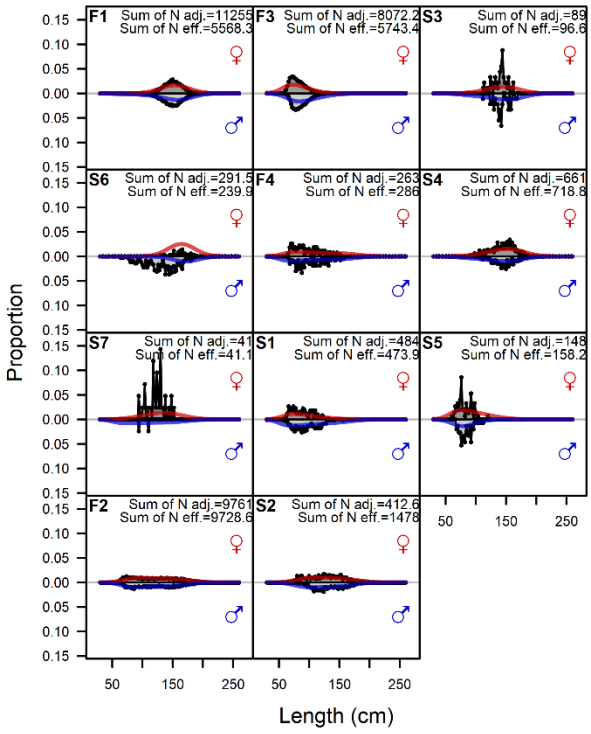


Figure 6. Example of the fleet-specific, time-aggregated length compositions from the DFA assessment of the first iteration of the SB29 simulation scenario for the sandbar shark as plotted using r4ss (Taylor *et al*., 2021).

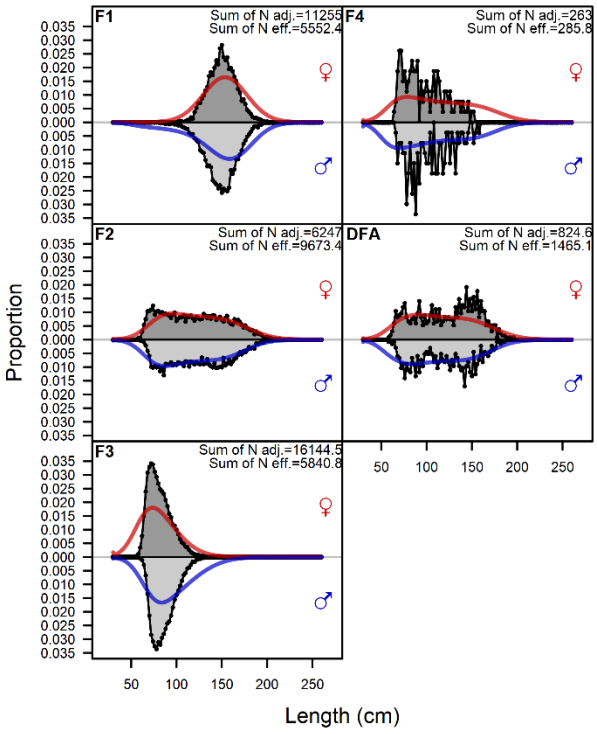


Figure 7. Estimated relative depletion (estimated depletion / simulated depletion) from CI assessments (darker shaded violins for each color; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins for each color; assessments fitted using DFA predicted trends as relative abundance inputs) in the Atlantic sharpnose shark simulations. Note that accurately estimated relative depletion should equal 1. Simulation scenarios are grouped based on trials in which survey catchability did not change ("no change"), survey catchability underwent a knife-edged shift in two surveys ("knife-edge"), and survey catchability underwent a gradual shift ("gradual"). Each row is separated based on the underlying fishing mortality (*F*) scenarios: no shift (Const *F*), increase in *F* (Inc *F*), decrease in *F* (Dec *F*), and an increase then decrease in *F* (U *F*). Simulations in which three surveys were simulated are in the left column and those for four surveys simulated are in the right column. Note the variable *y*-axes. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective *y*-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted. For full scenario-specific results, refer to the Supplementary material.

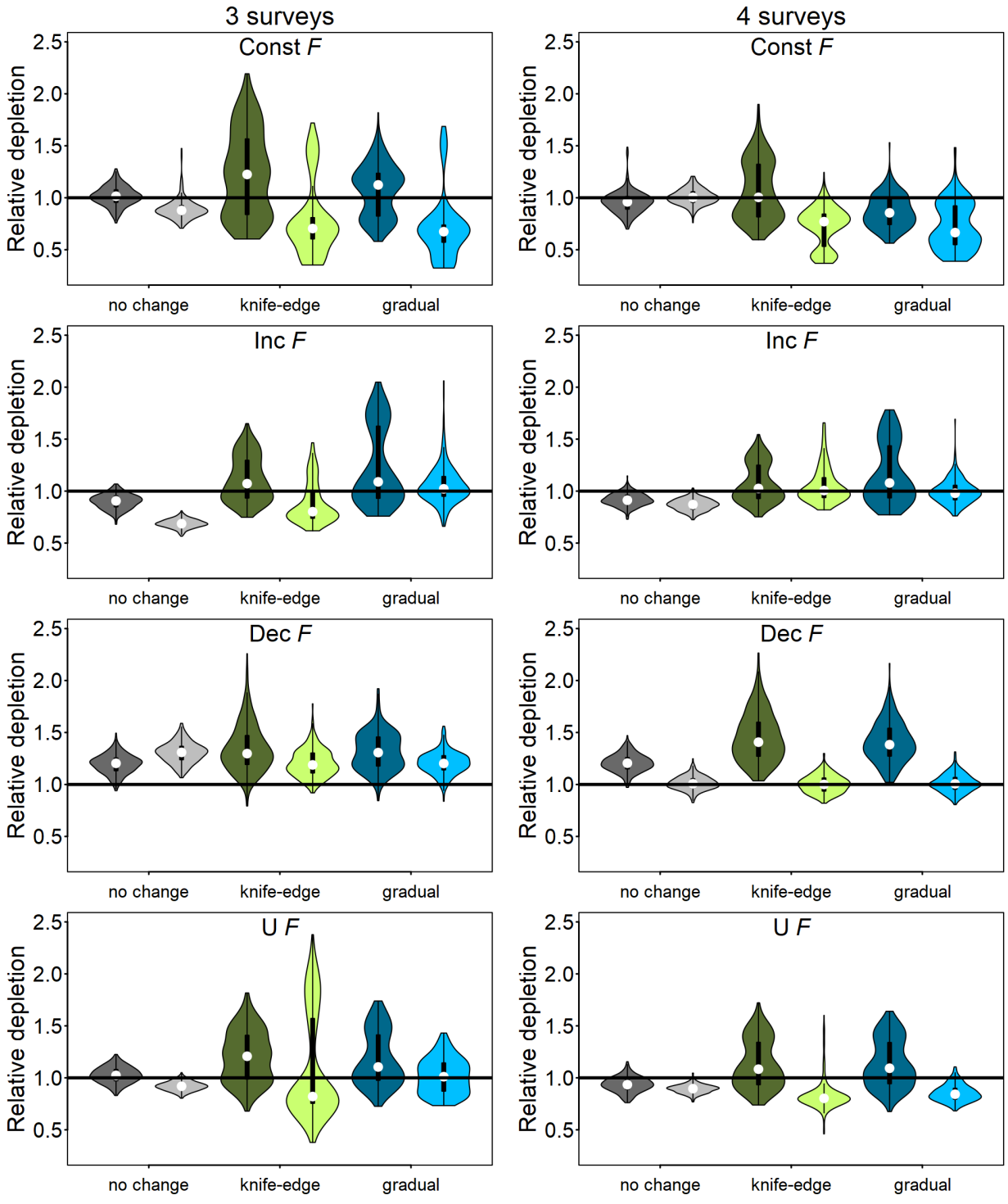

3 surveys
4 surveys
Const F
Inc F
Dec F
U F
Relative depletion
2.5
2.0
1.5
1.0
0.5
no change
knife-edge
gradual

Figure 8. Estimated relative final F (estimated F_{final} / simulated F_{final}) from CI assessments (darker shaded violins for each color; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins for each color; assessments fitted using DFA predicted trends as relative abundance inputs) in the Atlantic sharpnose shark simulations. Note that accurately estimated relative F_{final} should equal 1. Simulation scenarios are grouped based on trials in which survey catchability did not change ("no change"), survey catchability underwent a knife-edged shift in two surveys ("knife-edge"), and survey catchability underwent a gradual shift ("gradual"). Each row is separated based on the underlying fishing mortality (F) scenarios: no shift (Const F), increase in F (Inc F), decrease in F (Dec F), and an increase then decrease in F (U F). Simulations in which three surveys were simulated are in the left column and those four surveys simulated are in the right column. Note the variable y-axes. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted. For full scenario-specific results, refer to the Supplementary material.

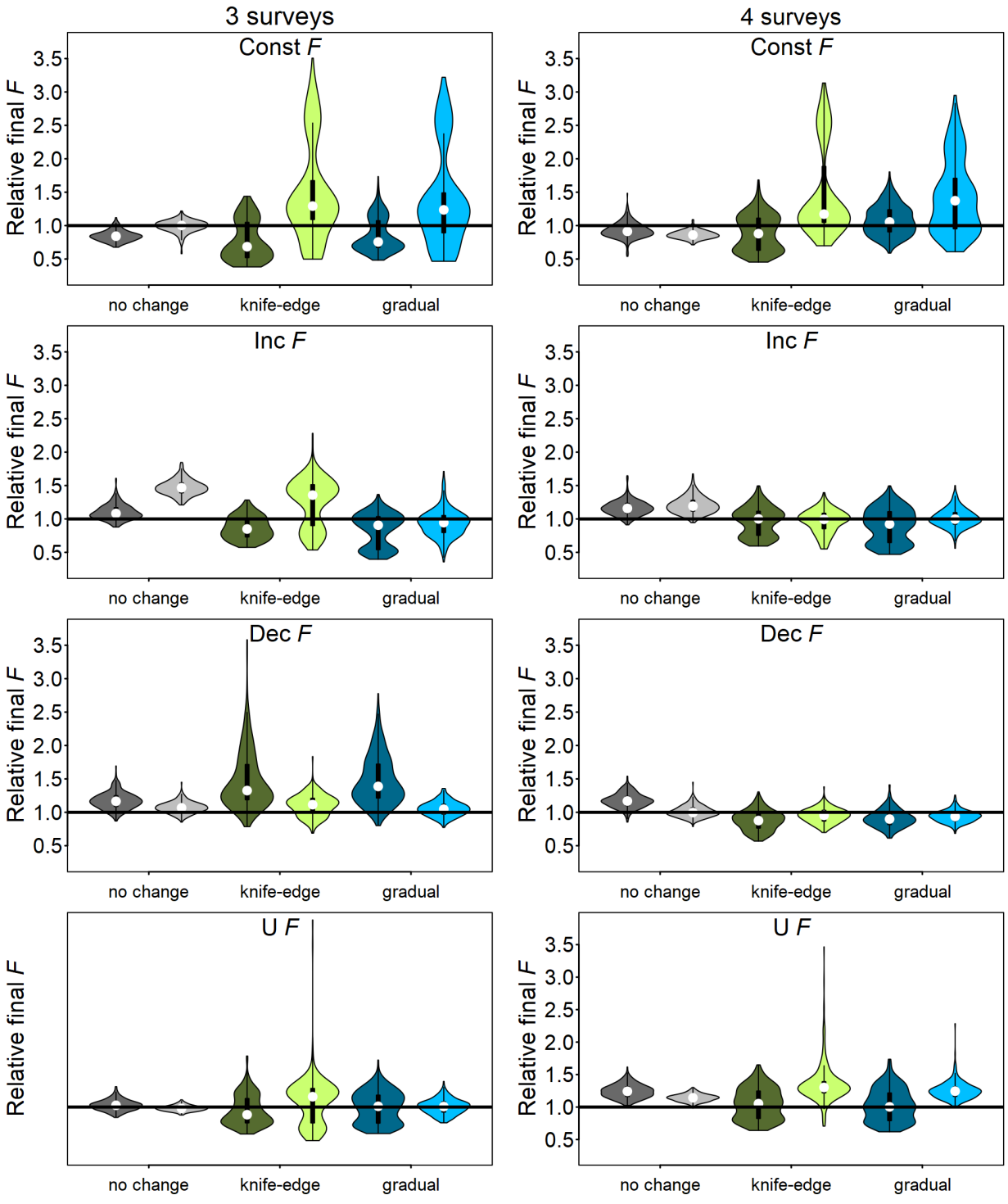

3 surveys
4 surveys
Const F
Inc F
Dec F
U F
Relative final F
3.5
3.0
2.5
2.0
1.5
1.0
0.5
no change
knife-edge
gradual

Figure 9. Estimated relative depletion (estimated depletion / simulated depletion) from CI assessments (darker shaded violins for each color; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins for each color; assessments fitted using DFA predicted trends as relative abundance inputs) in the sandbar shark simulations. Note that accurately estimated relative depletion should equal 1. Simulation scenarios are grouped based on the number of surveys that experienced shifting catchability (top row – 0 and 1, middle row – 2, bottom row – 3) and direction of shifting patterns in catchability ("up" denotes scenarios in which time-varying *q* increased for all surveys that experienced shifts; "down" denotes scenarios in which time-varying *q* decreased for all shifting surveys; "mix" denotes scenarios in which time-varying *q* increased for one or more surveys and decreased for one or more surveys; the first "mix" label (pink violins) denotes scenarios in which the first survey with time-varying *q* increased and the remaining surveys with time-varying *q* decreased, and the second "mix" label (purple violins) denotes scenarios in which the first survey with time-varying *q* underwent a decrease, and the remaining surveys with time-varying q underwent an increase). Simulations in which surveys with complete data were simulated are in the left column and those with missing data are in the right column. Note the variable *y*-axes. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective *y*-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted. For full scenario-specific results, refer to the Supplementary material.

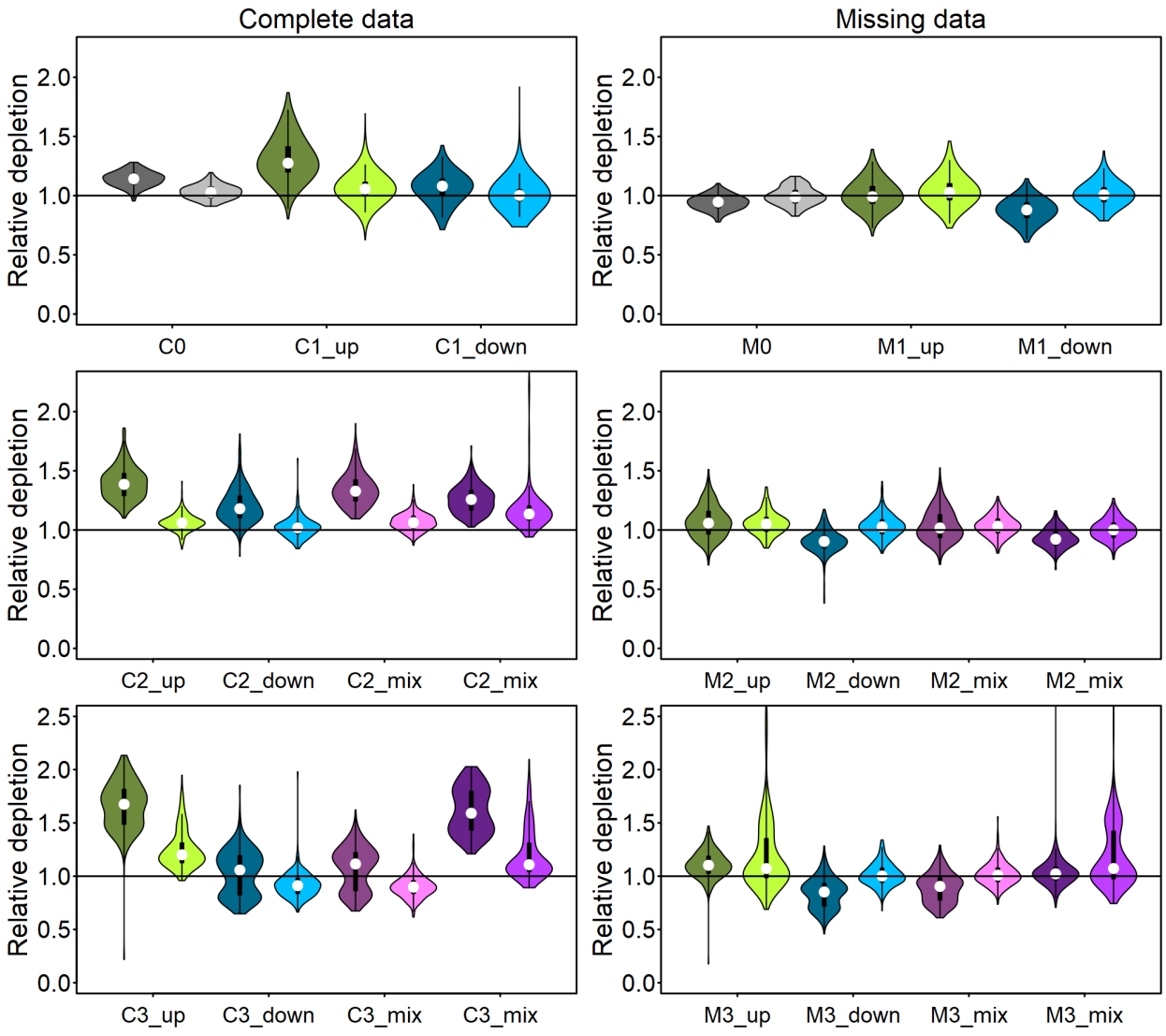
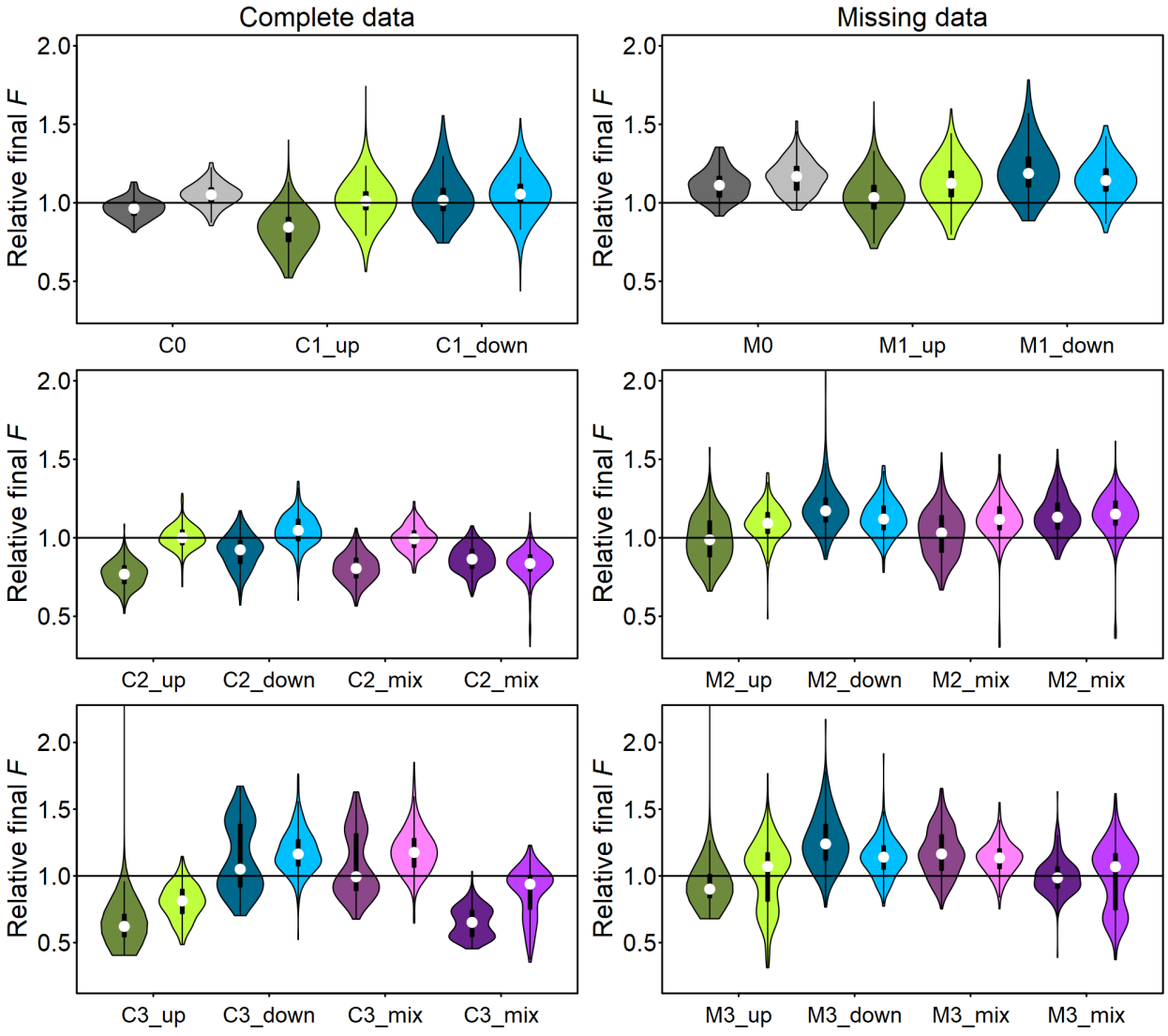

Complete data
Missing data
Relative depletion
0.0
0.5
1.0
1.5
2.0
2.5
C0
C1_up
C1_down
M0
M1_up
M1_down
C2_up
C2_down
C2_mix
C2_mix
M2_up
M2_down
M2_mix
M2_mix
C3_up
C3_down
C3_mix
C3_mix
M3_up
M3_down
M3_mix
M3_mix

Figure 10. Estimated relative final F (estimated F_{final} / simulated F_{final}) from CI assessments (darker shaded violins for each color; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins for each color; assessments fitted using DFA predicted trends as relative abundance inputs) in the sandbar shark simulations. Note that accurately estimated relative F_{final} should equal 1. Simulation scenarios are grouped based on the number of surveys that experienced shifting catchability (top row – 0 and 1, middle row – 2, bottom row – 3) and direction of shifting patterns in catchability ("up" denotes scenarios in which time-varying q increased for all surveys that experienced shifts; "down" denotes scenarios in which time-varying q decreased for all shifting surveys; "mix" denotes scenarios in which time-varying q increased for one or more survey and decreased for one or more surveys; the first "mix" label (pink violins) denotes scenarios in which the first survey with time-varying q increased and the remaining surveys with time-varying q decreased, and the second "mix" label (purple violins) denotes scenarios in which the first survey with time-varying q underwent a decrease, and the remaining surveys with time-varying q underwent an increase). Simulations in which surveys with complete data were simulated are in the left column and those with missing data are in the right column. Note the variable y-axes. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted. For full scenario-specific results, refer to the Supplementary material.


Complete data
Missing data
Relative final F
2.0
1.5
1.0
0.5
C0
C1_up
C1_down
M0
M1_up
M1_down
C2_up
C2_down
C2_mix
C2_mix
M2_up
M2_down
M2_mix
M2_mix
C3_up
C3_down
C3_mix
C3_mix
M3_up
M3_down
M3_mix
M3_mix

Supplementary material: ICES Journal of Marine Science, 78

Reconciling conflicting survey indices of abundance prior to stock assessment

Cassidy D. Peterson, Dean L. Courtney, Enric Cortés, and Robert J. Latour

Note: For full results detailing the ability of the DFA to reconcile multiple conflicting survey indices, including how closely DFA trends approximate simulated stock abundance and what sensitivities influenced survey weighting (e.g. as measured by factor loadings) and thereby DFA performance, we refer the reader to the companion manuscript (Peterson *et al*., 2021).

Dynamic factor analysis (DFA) rescaling approach from Peterson *et al*. (2021)

We developed a DFA rescaling approach to modify the input survey indices in a manner consistent with the requirements of DFA, which preserves the error structure and the relative scale of the survey indices, and allows for backtransformation of the resulting DFA index out of detrended, log-scale. The rescaling approach was simulation tested in Peterson *et al.* (2021) and is as follows:

1) Each survey index (i) was multiplied by a constant (c_i), which is comparable to redefining effort. The choice of c_i for each survey was determined iteratively, by arbitrarily defining a vector of all constants for each survey, $\boldsymbol{c} = [c_1, \ldots, c_n]$ for all indices i=1 to n, and adjusting each c_i until the conditions outlined below (Aa–Ad) were met within the steps of the rescaling approach outlined here.
2) All indices from step 1 were log-transformed, thereby normalizing survey error.
3) Each log-transformed survey index from step 2 was centered and demeaned by subtracting and dividing each index by the survey-specific mean.
4) The global standard deviation (*GSD*) was estimated for all demeaned survey indices, collectively (from step 3).
5) Each demeaned index was divided by the *GSD*, comparable to z-scoring the index.
6) The DFA was run.
7) The resulting DFA-predicted trend was back-transformed by first multiplying by the *GSD* and then exponentiating with bias correction. Annual standard errors estimated by the DFA model were multiplied by the *GSD* (following the transformation of variance rule: $SD(aX) = a \times SD(X)$, where a is a constant), representing lognormal error of the DFA trend.

A) The choice in c_i should result in the following conditions (a–c) being met within the steps of the rescaling approach outlined above:
    a. *The mean of each log-transformed survey index (from step 2) was greater than zero*. Note that if the logged survey-specific mean is close to zero, then the rescaling approach would not work, because we would essentially be dividing by zero in step 3. Further, if a survey index is relatively small, then the mean of the log transformed index may be negative, and dividing by a negative can reverse the trend of the

index. In practice, c_is that resulted in log-transformed survey index means ≳2.5, when possible, produced the best results.

b. *The resulting GSD (step 4) was small (GSD<<1).* When the GSD is greater than one, the scale of the back-transformed DFA-predicted trend is affected, resulting in unrealistic predicted changes in abundance.

c. *Most importantly, the standard deviation of each resulting transformed index (step 5) was approximately equal to 1.* This ensures that the format of the input survey index most appropriately approximates a z-scored index, with a mean of zero and a standard deviation of 1, as recommended in DFA applications (Holmes et al. 2014). When $\boldsymbol{c}$ was chosen so that the standard deviation of an input index was greater than one, the resulting DFA-predicted trend overfitted the corresponding survey, and fit more poorly to the remaining surveys.

d. *The resulting back-transformed trend should follow changes in magnitude consistent with those of the survey inputs.* Multiple combinations of $\boldsymbol{c}$ can fit the above requirements; however, we want to choose constants that preserve the relative trend of the input surveys. For example, if the starting and ending points of a (or multiple) largely reliable input survey(s) change by an order of magnitude of ~2, then the resulting back-transformed trend may not be appropriate if it changes by an order of 0.5 or 4. Try to maintain a consistent and reliable trend following the raw survey indices. This is obviously challenging if the analyst is unsure of what survey inputs are reliable and the user may have to rely on best judgement. Note that this recommendation is not important if the magnitude of change of the resulting DFA index is not of interest and if the underlying abundance trend is not a one-way trip.

We recommend ensuring that the raw survey index follows the same general pattern across each step of the rescaling process, and that the resulting trend is realistic given the input data. In our application, the pattern of the DFA trend estimated from our rescaling approach was very similar to the DFA trend estimated from a log-transformed, then traditionally z-scored survey index. (Although, in a traditionally run DFA, we cannot back transform the resulting DFA trend out of log-space). We also recommend ensuring that the mean fit ratio is low, as the mean fit ratio will increase with a less ideal vector of constants, $\boldsymbol{c}$.

It is important to note that the ability of the DFA assessment to accurately estimate relative depletion and fishing mortality in the final year of the EM (F_{final}) depends on the appropriate implementation of the rescaling approach. In the Atlantic sharpnose shark simulation, the choice in constants in the DFA rescaling protocol affected the results of the corresponding assessment. These effects were greater in the 'one-way trip' scenarios (F2 and F3), because the abundance pattern was not 'bounded' as in more complex fishing mortality scenarios (F4). The vector of constants was less important in the sandbar shark scenario, likely because the trend was slightly more complex than a 'one-way trip' and because the greater number of surveys reduced the possible combinations of constants that could be applied to each scenario.

Treatment of length composition data for the DFA estimation model

For the DFA EM, length compositions were generated by first tabulating the number of observations that fell within each length bin for each survey, as is the traditional approach for calculating length composition data within Stock Synthesis. These length compositions for each survey were then multiplied by a weighting factor. The weighting factor was obtained by calculating the relative strength of each factor loading (Γ_j/max(Γ)). Consider the first iteration of the Atlantic sharpnose simulation scenario UF_knife_1:

1. Generate length compositions for each survey (LenComp_1, LenComp_2, Lencomp_3)
2. Get factor loadings for each survey: Γ_1=0.39, Γ_2=0.02, Γ_3=0.38.
3. Calculate factor loading based weighting factor ($w_j = \Gamma_j/\max(\boldsymbol{\Gamma})$): w_1=1.00, w_2=0.05, w_3=0.97.
4. Multiply the corresponding length composition with the weighting factor and sum across surveys to obtain the DFA length composition: $\text{LenComp}_{\text{DFA}} = \Sigma_j(\text{LenComp}_j \times w_j)$

We set the weight for negative factor loadings equal to zero, though a small constant may also be explored as an appropriate treatment.

Alternate axes of uncertainty explored

The Atlantic sharpnose shark simulation study also explored the impacts of (1) survey variability (via coefficients of variation; CV) and (2) number of surveys.

Survey variability

The CI assessment performance varied slightly based on survey *CV* (e.g. the difference between trials C22–C24; Figures S2–S5) and depended on the combination of survey variability and shift in catchability. Surveys for which *CV* was smaller were more heavily weighted in the DFA trend (Peterson *et al*., 2021), which was also reflected in assessment performance estimates. As noted in the main text, when surveys with the smallest *CV*s underwent time-varying shifts in catchability in the opposite direction of the predominant trend in stock abundance, estimated depletion and terminal *F* were more biased in the DFA assessment (i.e., increasing catchability when the stock size was decreasing; IncF_k2, IncF_g2, ConstF_k2, constF_g2). Otherwise, the DFA assessment results were generally more consistent and precise across variations in survey variability compared to CI assessment results. Likewise, several CI assessment scenarios also experienced biases when catchability shifted in surveys with low *CV*s (e.g. DecF_k2, DecF_g2, SB C105, etc.; Figures S2–S9).

Number of surveys

When an incomplete fourth survey index was added to each assessment model, the accuracy and precision of estimated depletion and F_{final} generally improved in both the CI and DFA assessment models (except in F1; Figures 1–2; e.g. consider C22–24 compared to D22–24; Figures S2–S5).

Management reference points

Across all simulations for each species, relevant management quantities, such as fishing mortality that would produce maximum sustainable yield, MSY, (F_{MSY}; Figure S9), F_{ratio} (terminal F/F_{MSY}; Figure S10), spawning stock biomass that produces MSY (SSB_{MSY}), SSB_{ratio} (terminal SSB/SSB_0; Figure S12), and the ratio of SSB_{MSY} to virgin spawning stock biomass (SSB_{MSY}/SSB_0; Figure S13) were examined. Notably, variations in key reference point estimates tend to follow observed patterns in corresponding parameter estimates (e.g. estimates in F_{ratio} tended to correspond to bias

patterns in F_{final} and estimates of SSB_{MSY} tended to follow biases in relative depletion; compare plots –S10 and S2–S12). There were distinct differences in estimated F_{MSY} for DecF and U F fishing mortality scenarios for the Atlantic sharpnose shark (Figure S9), though these differences were mostly eliminated when calculating F_{ratio} (Figure S10). Estimates of SSB_{MSY} were similar across EMs (Figure S11).

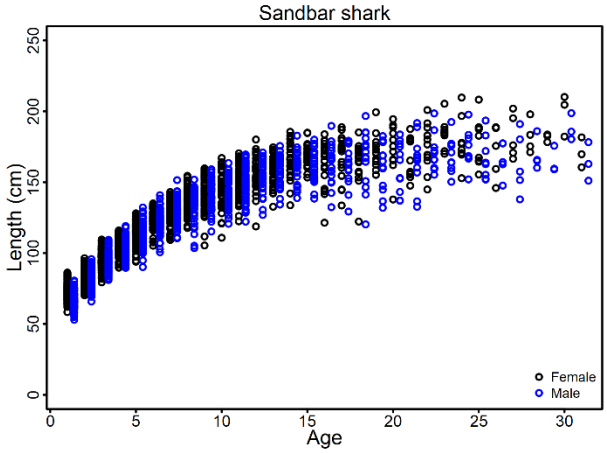


Figure S1. Example of the sex-specific, age-length relationship for the sandbar shark as obtained from the terminal year of the simulation Trial SB101.

Atlantic sharpnose shark complete results

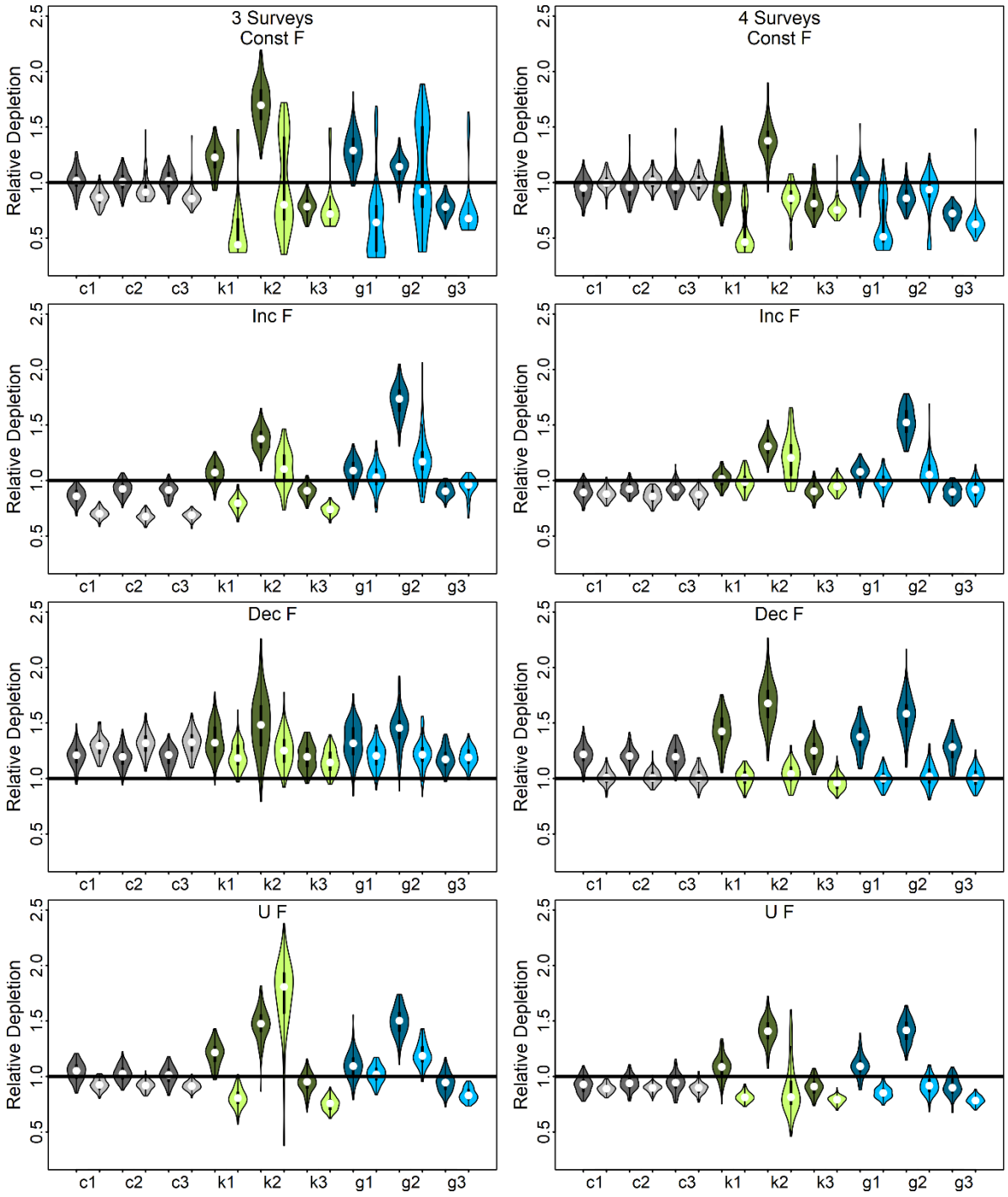


Figure S1. Estimated relative depletion (estimated depletion / simulated depletion) from CI assessments (darker shaded violins; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins; assessments fitted using DFA predicted trends as relative abundance inputs) in the Atlantic sharpnose shark simulations. Note that

accurately estimated relative depletion should equal 1. Simulation scenarios correspond to Table 2, where c, k, and g indicate 'constant,' 'knife-edged,' and 'gradual' shifts in catchability. Each row is separated based on the underlying fishing mortality (F) scenarios: no shift (F1), increase in F (F2), decrease in F (F3), and an increase then decrease in F (F4). Simulations in which three surveys were simulated are in the left column and four surveys simulated are in the right column. Note the variable y-axes. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted.

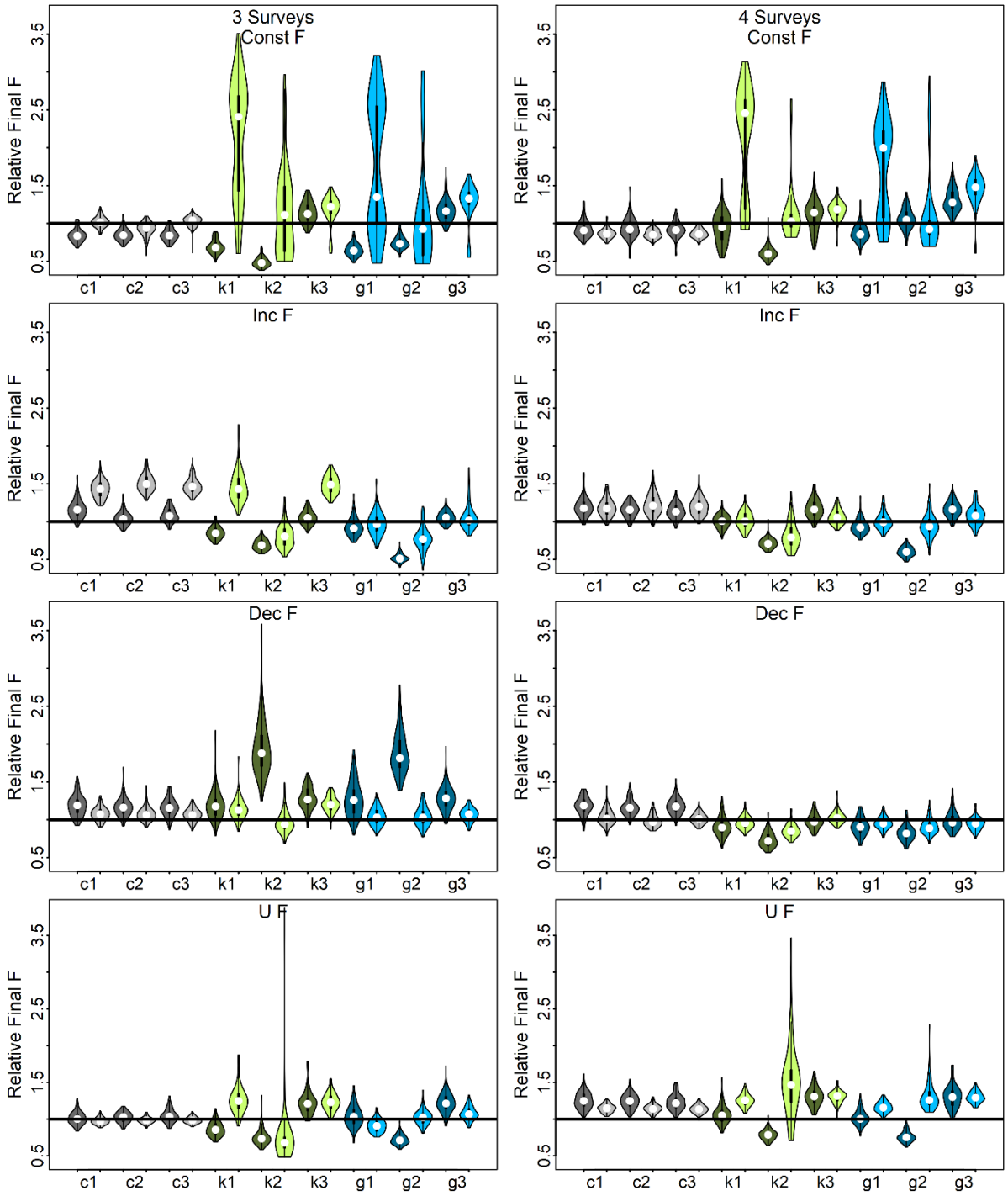


Figure S2. Estimated relative final F (estimated F_{final} / simulated F_{final}) from CI assessments (darker shaded violins; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins; assessments fitted using DFA predicted trends as relative abundance inputs) in the Atlantic sharpnose shark simulations. Note that accurately estimated relative depletion should equal 1. Simulation scenarios correspond to Table 2, where c, k, and g indicate 'constant,' 'knife-edged,' and 'gradual' shifts in catchability. Each row is separated based on the underlying fishing mortality (F) scenarios: no shift (F1), increase in F (F2),

decrease in F (F3), and an increase then decrease in F (F4). Simulations in which three surveys were simulated are in the left column and four surveys simulated are in the right column. Note the variable y-axes. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted.

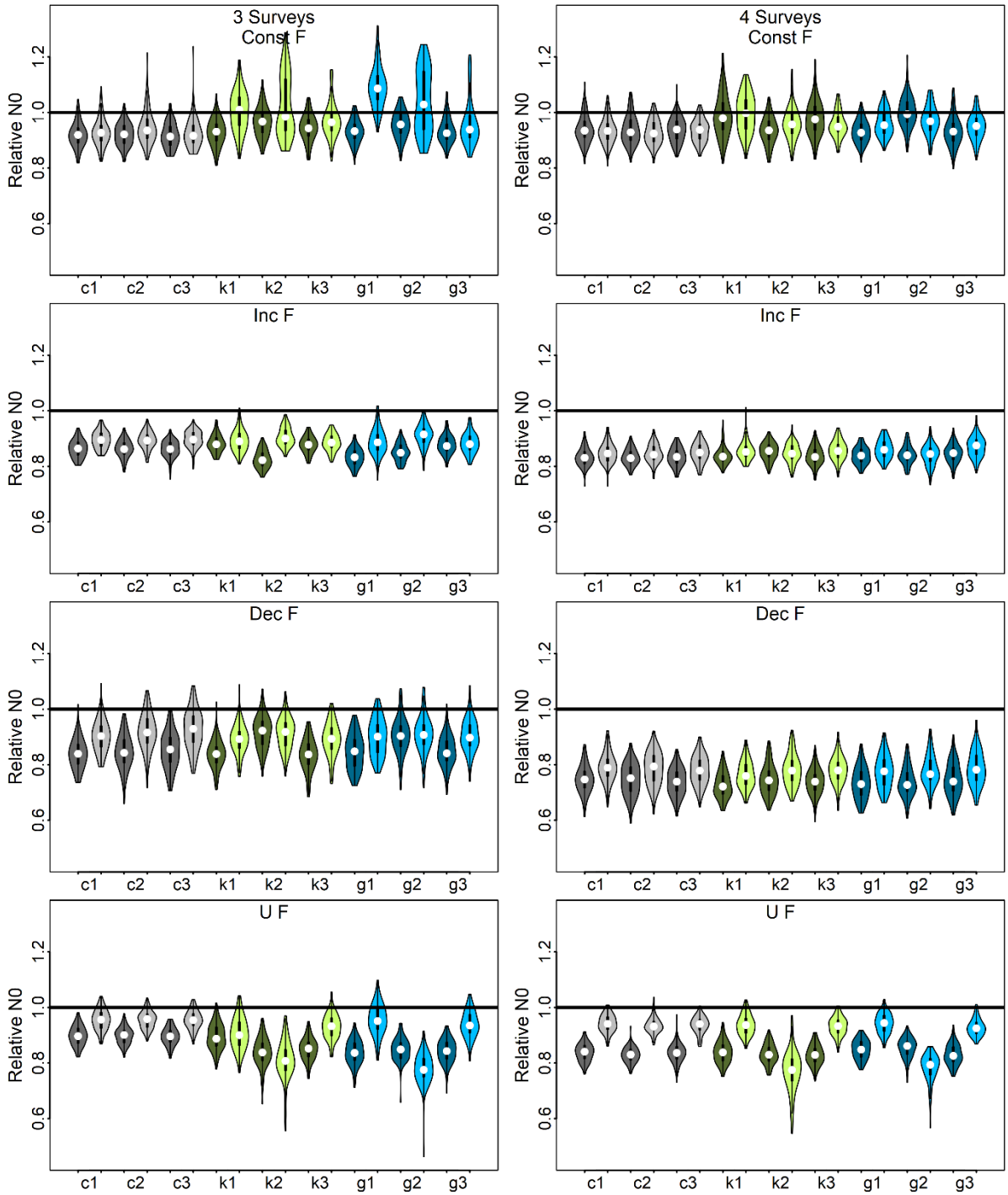


Figure S3. Estimated relative virgin abundance in numbers (estimated N_0 / simulated N_0) from CI assessments (darker shaded violins; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins; assessments fitted using DFA predicted trends as relative abundance inputs) in the Atlantic sharpnose shark simulations. Note that accurately estimated relative N_0 should equal 1. Simulation scenarios correspond to Table 2, where c, k, and g indicate 'constant,' 'knife-edged,' and 'gradual' shifts in catchability. Each row is separated based on the underlying fishing mortality (F) scenarios: no shift (F1), increase in F

(F2), decrease in F (F3), and an increase then decrease in F (F4). Simulations in which three surveys were simulated are in the left column and four surveys simulated are in the right column. Note the variable y-axes. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted.

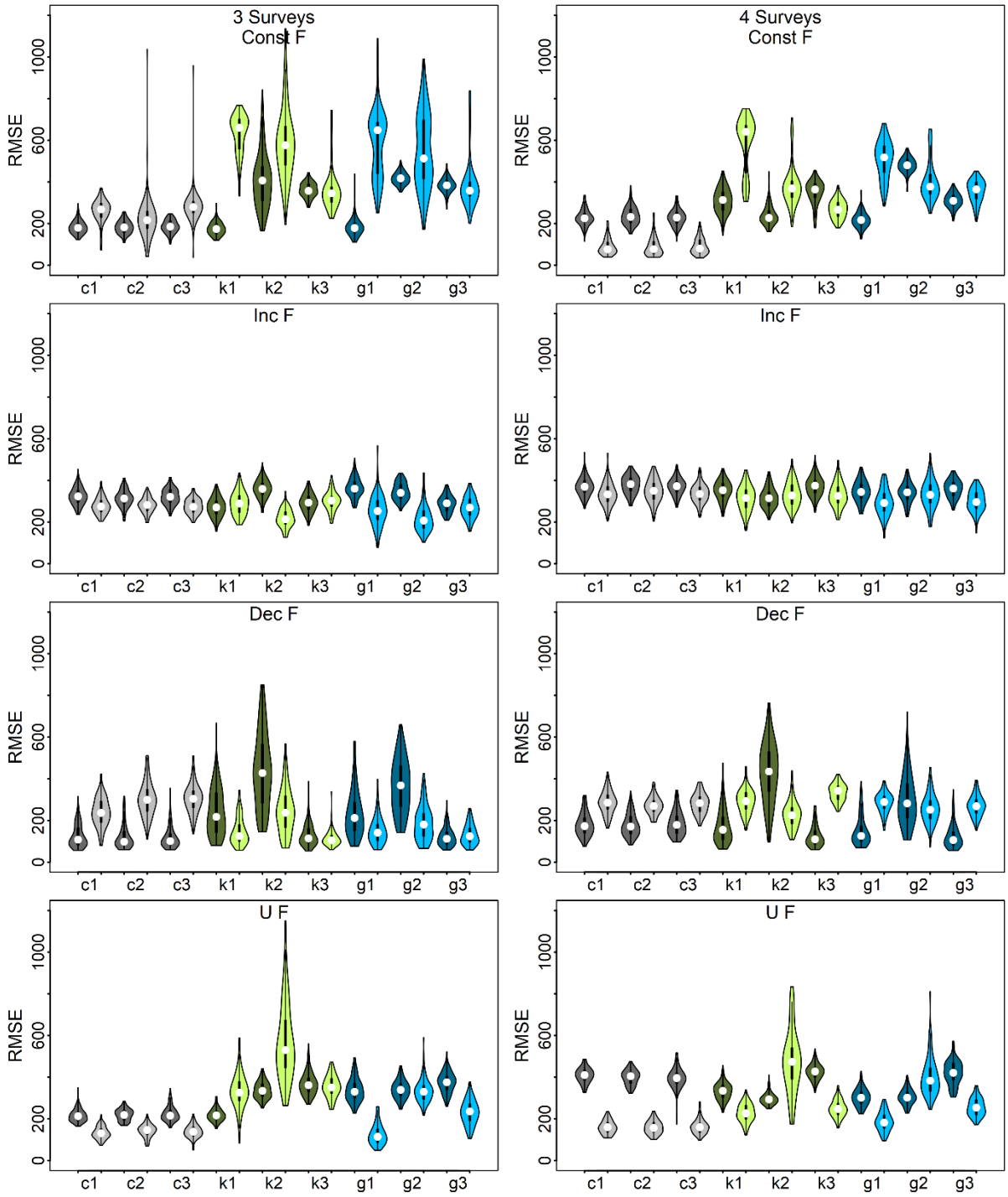


Figure S4. Estimated root-mean square error (RMSE) of predicted abundance in numbers of fish from CI assessments (darker shaded violins; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins; assessments fitted using DFA predicted trends as relative abundance inputs) in the Atlantic sharpnose shark simulations. Lower RMSE indicates a more accurately estimated abundance series. Simulation scenarios correspond to Table 2, where c, k, and g indicate 'constant,' 'knife-edged,' and 'gradual' shifts in catchability. Each row is separated based on the underlying fishing mortality (F) scenarios: no

shift (F1), increase in F (F2), decrease in F (F3), and an increase then decrease in F (F4). Simulations in which three surveys were simulated are in the left column and four surveys simulated are in the right column. Note the variable y-axes. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted.

Sandbar shark complete results

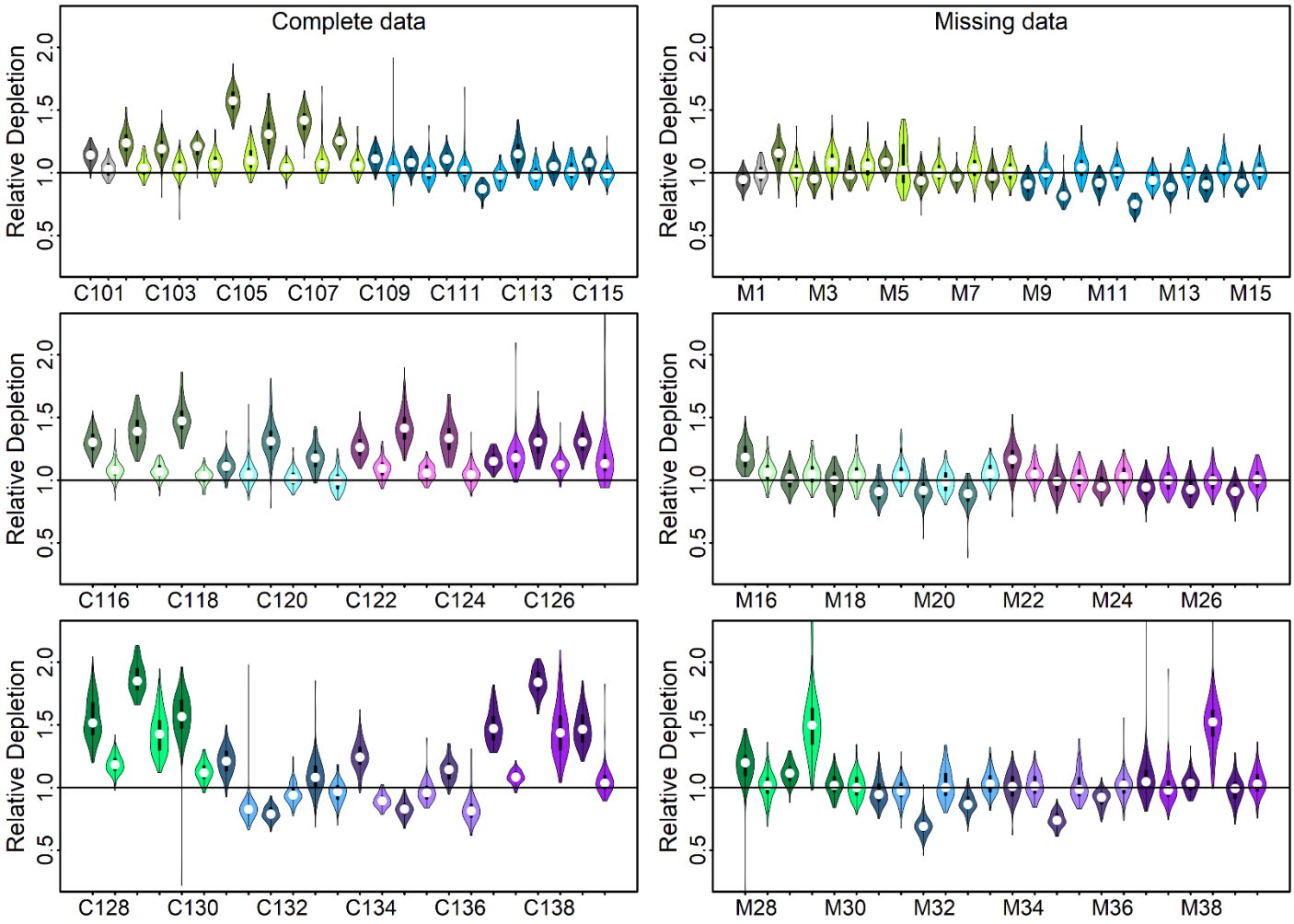


Figure S5. Estimated relative depletion (estimated depletion / simulated depletion) from CI assessments (darker shaded violins; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins; assessments fitted using DFA predicted trends as relative abundance inputs) in the sandbar shark simulations. Note that accurately estimated relative depletion should equal 1. Simulation scenarios correspond to Table 3. Simulations in which surveys with complete data were simulated are in the left column and with missing data are in the right column. Note the variable y-axes. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted.

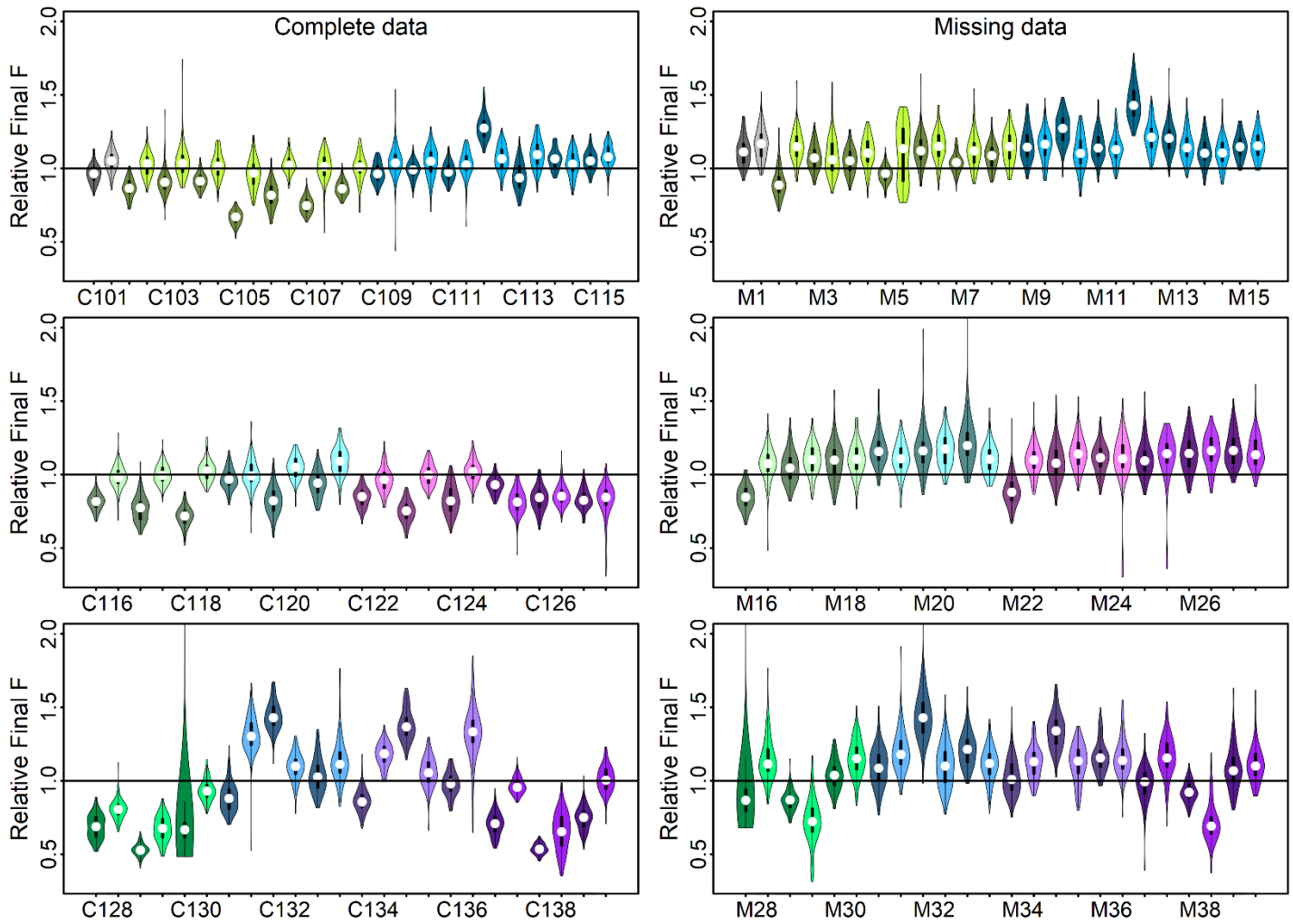


Figure S6. Estimated relative final F (estimated F_{final} / simulated F_{final}) from CI assessments (darker shaded violins; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins; assessments fitted using DFA predicted trends as relative abundance inputs) in the sandbar shark simulations. Note that accurately estimated relative F_{final} should equal 1. Simulation scenarios correspond to Table 3. Simulations in which surveys with complete data were simulated are in the left column and with missing data are in the right column. Note the variable y-axes. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted.

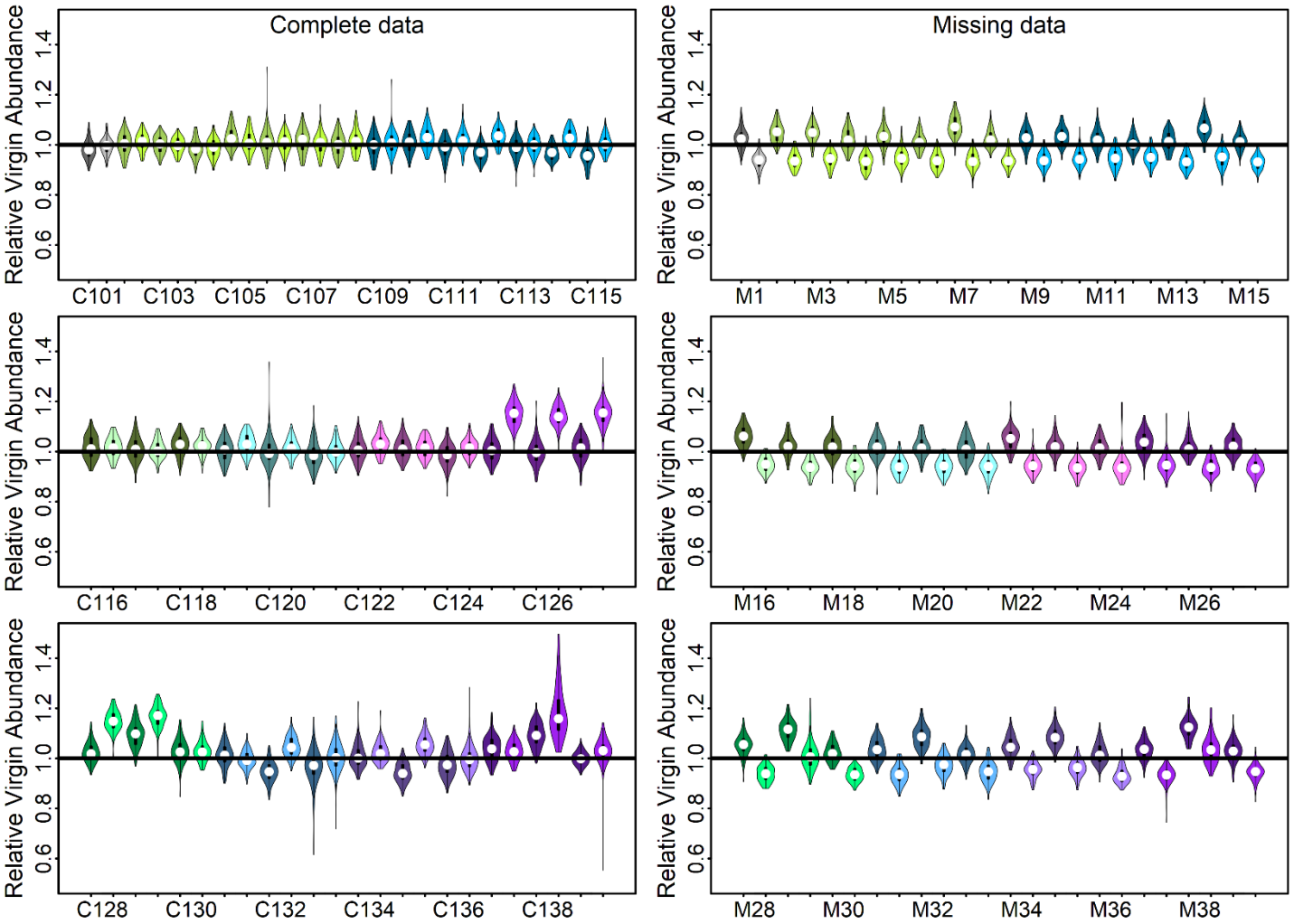


Figure S7. Estimated relative virgin abundance in numbers of fish (estimated N_0 / simulated N_0) from CI assessments (darker shaded violins; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins; assessments fitted using DFA predicted trends as relative abundance inputs) in the sandbar shark simulations. Note that accurately estimated relative N_0 should equal 1. Simulation scenarios correspond to Table 3. Simulations in which surveys with complete data were simulated are in the left column and with missing data are in the right column. Note the variable y-axes. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted.

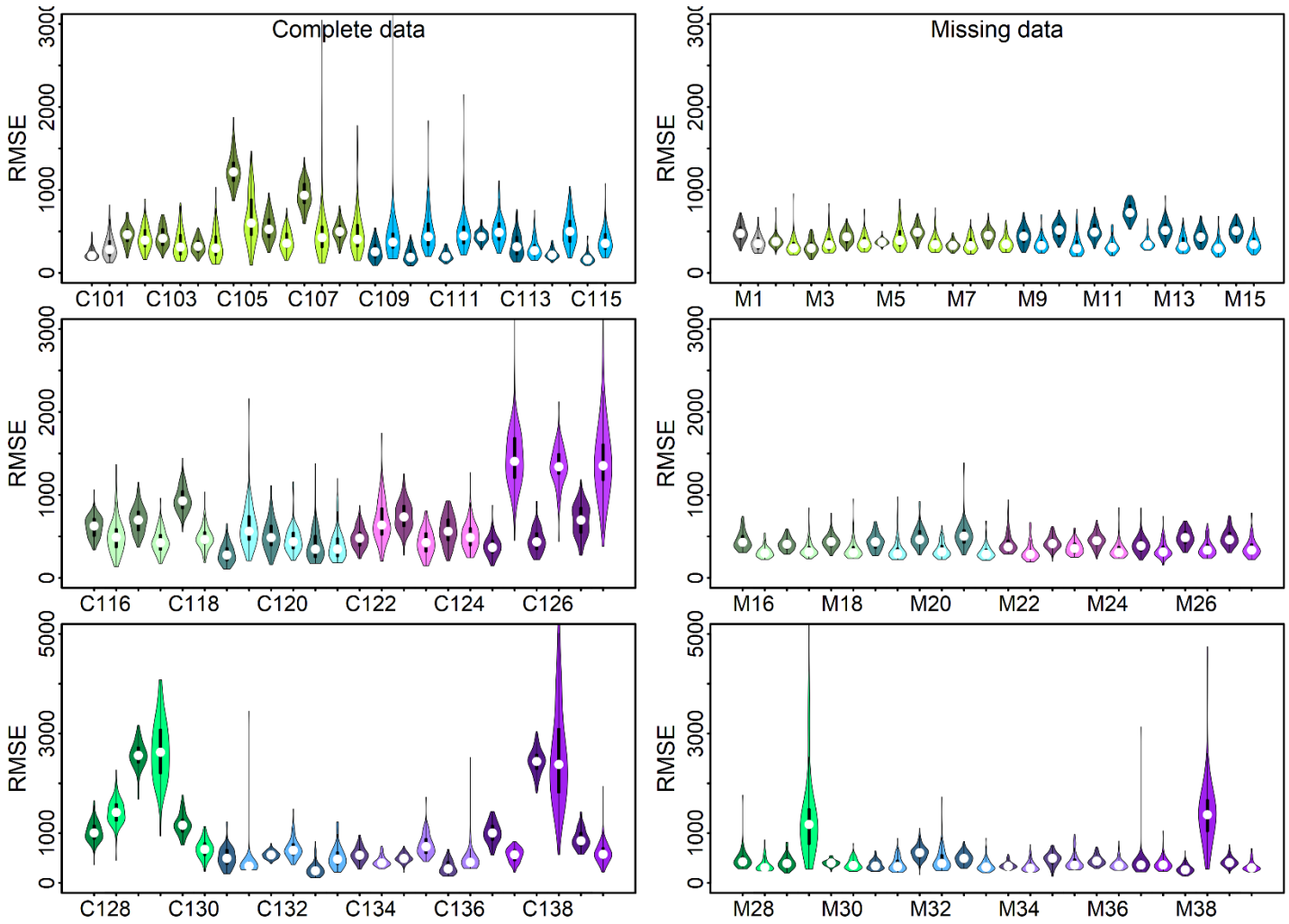


Figure S8. Estimated root-mean square error (RMSE) of predicted abundance in numbers of fish from CI assessments (darker shaded violins; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins; assessments fitted using DFA predicted trends as relative abundance inputs) in the sandbar shark simulations. Lower RMSE indicates a more accurately estimated abundance series. Simulation scenarios correspond to Table 3. Simulations in which surveys with complete data were simulated are in the left column and with missing data are in the right column. Note the variable y-axes. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted.

Atlantic sharpnose shark management reference point results

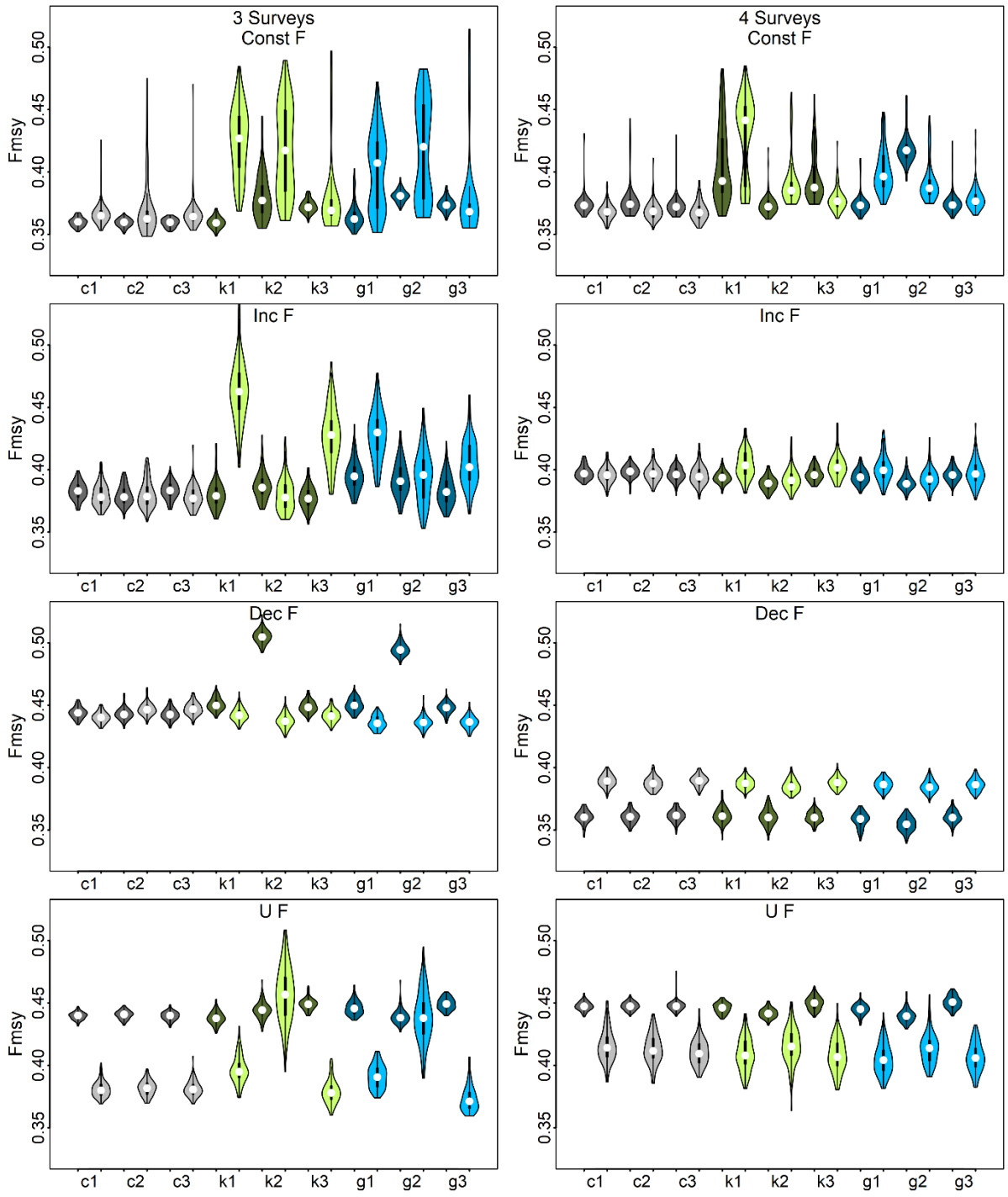


Figure S9. Estimated F_{MSY} from CI assessments (darker shaded violins; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins; assessments fitted using DFA predicted trends as relative abundance inputs) in the Atlantic sharpnose shark

simulations. Simulation scenarios correspond to Table 2, where c, k, and g indicate 'constant,' 'knife-edged,' and 'gradual' shifts in catchability. Each row is separated based on the underlying fishing mortality (F) scenarios: no shift (F1), increase in F (F2), decrease in F (F3), and an increase then decrease in F (F4). Simulations in which three surveys were simulated are in the left column and four surveys simulated are in the right column. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted.

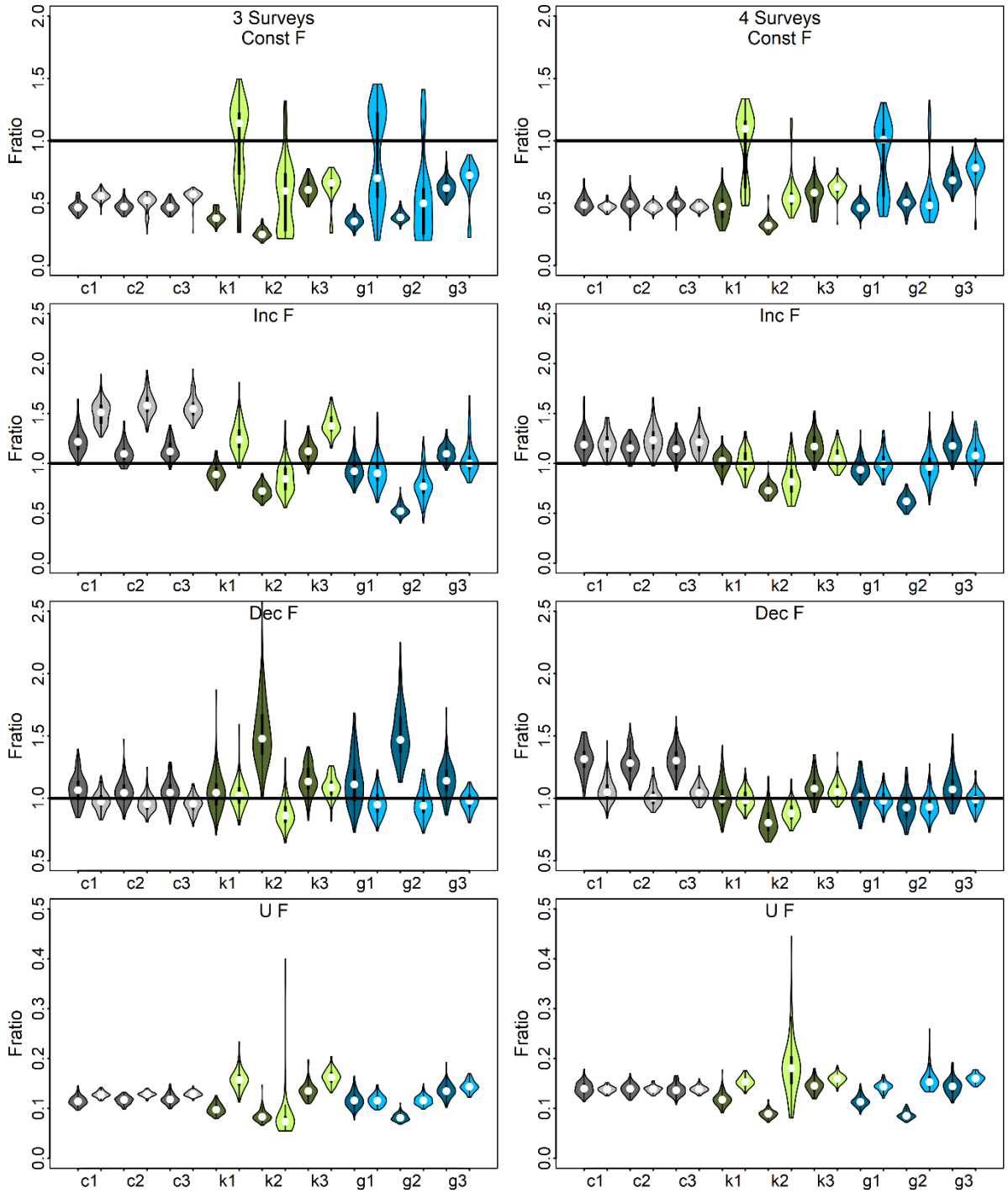


Figure S10. Estimated F_{ratio} (F_{final}/F_{MSY}) from CI assessments (darker shaded violins; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins; assessments fitted using DFA predicted trends as relative abundance inputs) in the Atlantic sharpnose shark simulations. Simulation scenarios correspond to Table 2, where c, k, and g indicate 'constant,' 'knife-edged,' and 'gradual' shifts in catchability. Each row is separated based on the underlying fishing mortality (F) scenarios: no shift (F1), increase in F (F2), decrease in F (F3), and an increase then decrease in F (F4). Simulations in which three

surveys were simulated are in the left column and four surveys simulated are in the right column. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted.

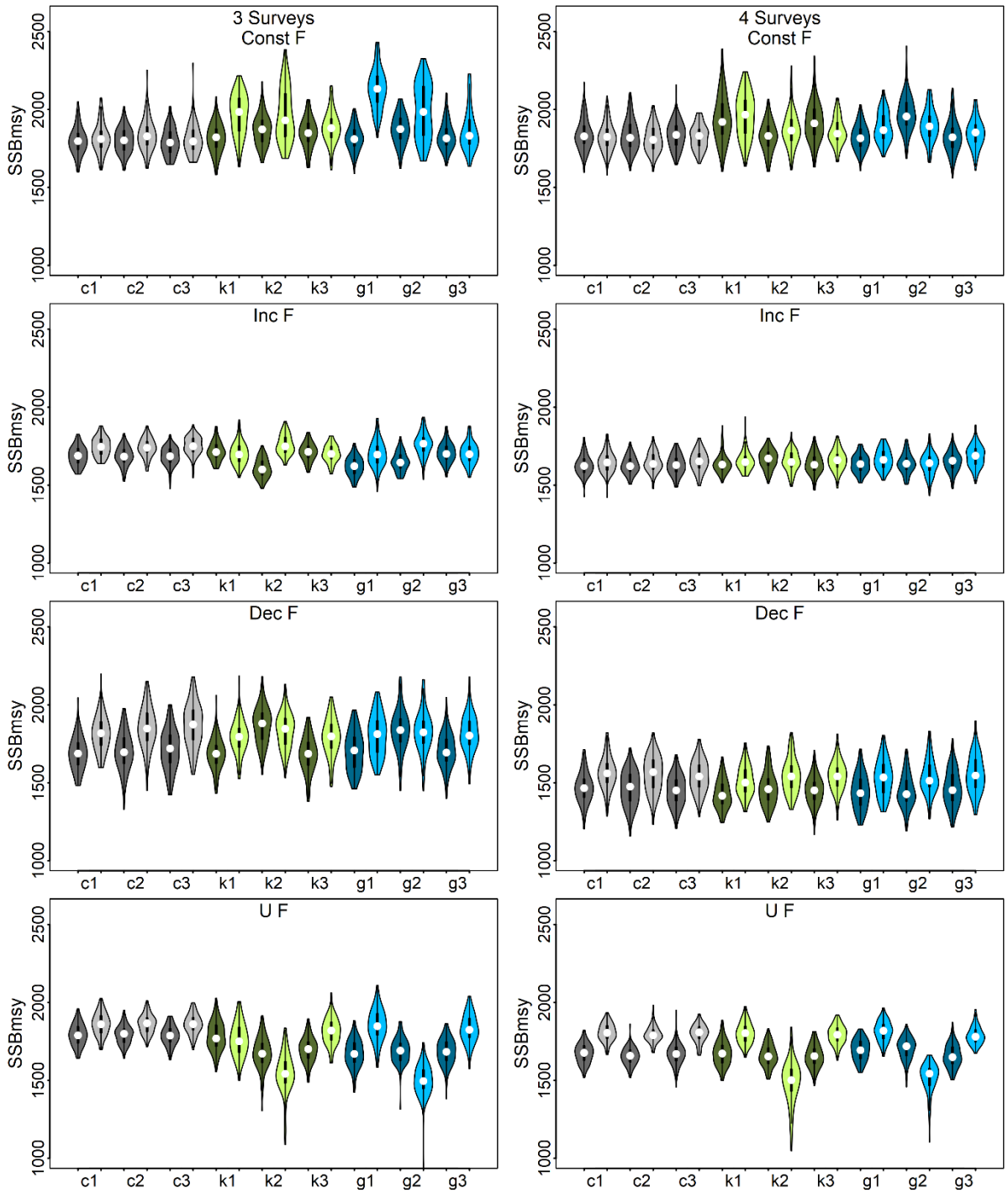


Figure S11. Estimated SSB_{MSY} (spawning-stock biomass that supports maximum sustainable yield) from CI assessments (darker shaded violins; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins; assessments fitted using DFA predicted trends as relative abundance inputs) in the Atlantic sharpnose shark simulations. Simulation scenarios correspond to Table 2, where c, k, and g indicate 'constant,' 'knife-edged,' and 'gradual' shifts in catchability. Each row is separated based on the underlying fishing mortality (F) scenarios: no shift (F1), increase in F (F2), decrease in F (F3), and an increase then

decrease in F (F4). Simulations in which three surveys were simulated are in the left column and four surveys simulated are in the right column. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted.

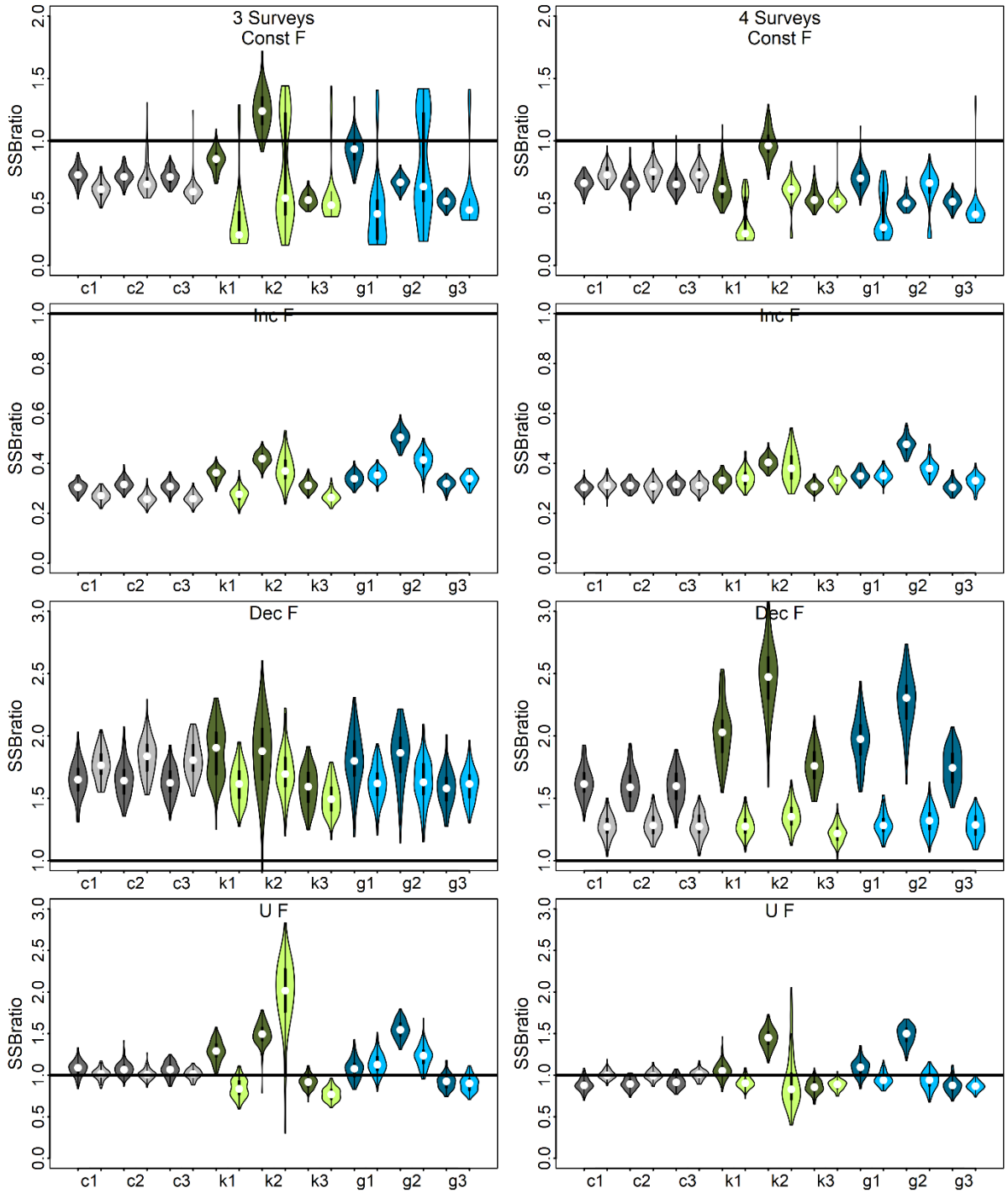


Figure S12. Estimated SSB_{ratio} (SSB_{final}/SSB_0) from CI assessments (darker shaded violins; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins; assessments fitted using DFA predicted trends as relative abundance inputs) in the Atlantic sharpnose shark simulations. Simulation scenarios correspond to Table 2, where c, k, and g indicate 'constant,' 'knife-edged,' and 'gradual' shifts in catchability. Each row is separated based on the underlying fishing mortality (F) scenarios: no shift (F1), increase in F (F2), decrease in F (F3), and an increase then decrease in F (F4). Simulations in which three

surveys were simulated are in the left column and four surveys simulated are in the right column. The shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted.

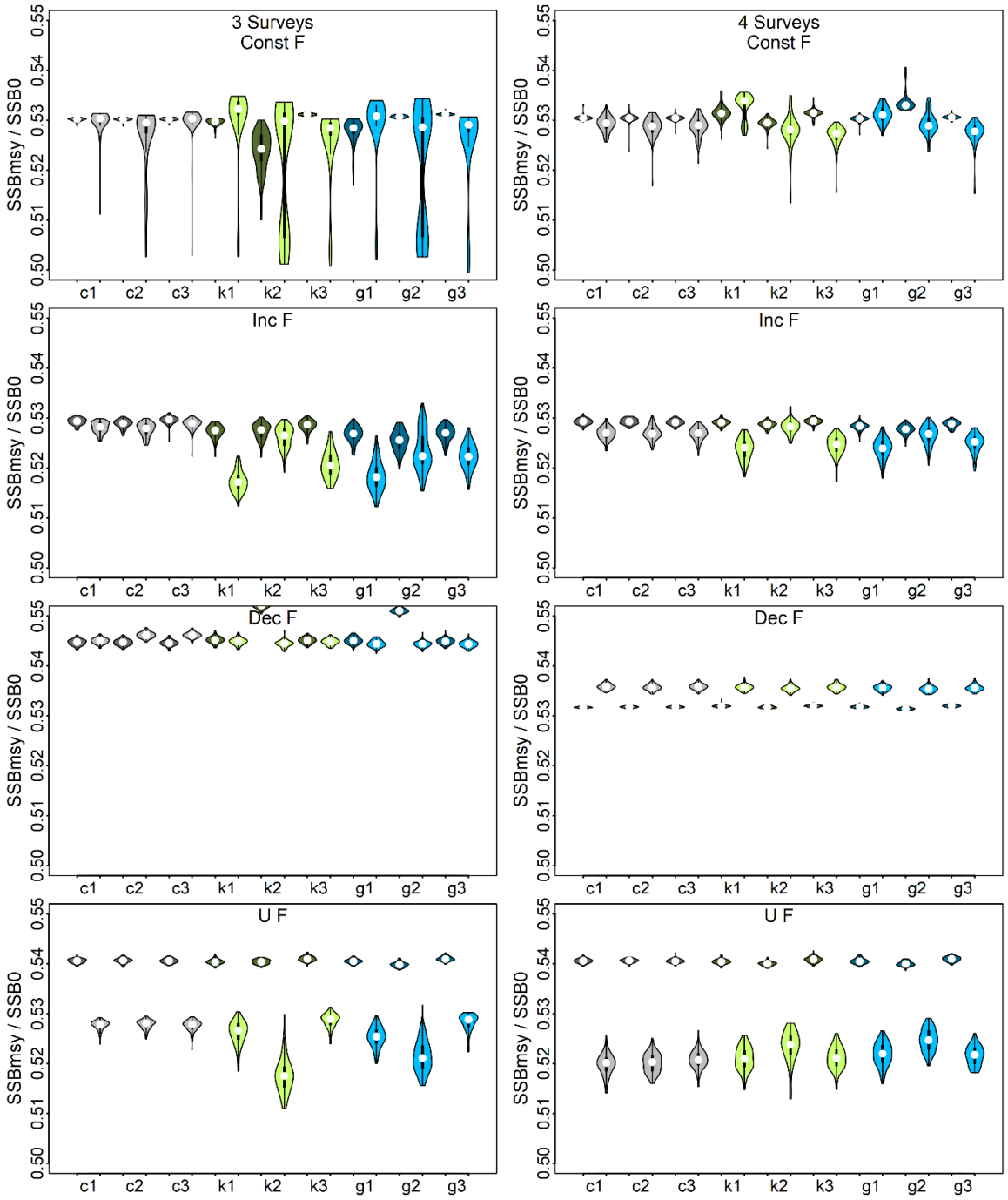


Figure S13. Estimated SSB_{MSY}/SSB_0 from CI assessments (darker shaded violins; assessments fitted using conflicting survey indices) compared to DFA assessments (lighter shaded violins; assessments fitted using DFA predicted trends as relative abundance inputs) in the Atlantic sharpnose shark simulations. Simulation scenarios correspond to Table 2, where c, k, and g indicate 'constant,' 'knife-edged,' and 'gradual' shifts in catchability. Each row is separated based on the underlying fishing mortality (F) scenarios: no shift (F1), increase in F (F2), decrease in F (F3), and an increase then decrease in F (F4). Simulations in which three surveys were simulated are in the left column and four surveys simulated are in the right column. The

shape of the violins indicate the distribution of parameter estimates, where the width of the violin corresponds to the quantity of results that fall at that respective y-value. The median (white dot), interquartile range (black bar), and upper/lower adjacent values corresponding to a box plot limits (thin black line) are also highlighted.

Table S1. List of age-specific proportion mature, natural mortality, and female fecundity for the simulated Atlantic sharpnose shark from SEDAR (2013).

Age	M	Proportion mature	Fecundity (no. female pups)
1	0.209	0.185	0.401
2	0.209	0.953	0.762
3	0.209	0.999	1.133
4	0.209	1.000	1.448
5	0.209	1.000	1.686
6	0.209	1.000	1.852
7	0.209	1.000	1.963
8	0.209	1.000	2.035
9	0.209	1.000	2.081
10	0.209	1.000	2.110
11	0.209	1.000	2.128
12	0.209	1.000	2.139
13	0.209	1.000	2.146
14	0.209	1.000	2.150
15	0.209	1.000	2.153
16	0.209	1.000	2.155
17	0.209	1.000	2.156
18	0.209	1.000	2.156

Table S2. List of parameter values in the Atlantic sharpnose shark operating model. Steepness (h) and von Bertalanffy growth parameters were taken from SEDAR (2013).

Parameter	Value
Survival-based stock–recruitment	
Z_0	1.9044
Z_{min}	0.209
S_{max}	0.8114
S_0	0.1489
Z_{frac}	0.8903
β	0.5808
h	0.56
$log(R_0)$	6.2035
R_0	494.4797
$Npups_0$	3320.671
S_{frac}	0.89
von Bertalanffy growth	
K_{avg}	0.49
$L_{\infty\ avg}$	81.6
t_0	–0.97

Table S3. List of age-specific female proportion mature, female natural mortality, and fecundity for the simulated sandbar shark population from SEDAR (2017).

Age	Natural mortality (M)	Proportion mature	Fecundity (no. pups)
0	0.1604	0	6.1390
1	0.1604	0	6.6107
2	0.1604	0	7.0274
3	0.1604	0	7.3955
4	0.1604	0	7.7206
5	0.1604	0	8.0079
6	0.1578	0.01	8.2616
7	0.1168	0.02	8.4857
8	0.1168	0.03	8.6838
9	0.1168	0.06	8.8587
10	0.1168	0.12	9.0132
11	0.1168	0.21	9.1497
12	0.1168	0.33	9.2703
13	0.1168	0.49	9.3768
14	0.1168	0.65	9.4709
15	0.1168	0.78	9.5540
16	0.1168	0.88	9.6274
17	0.1168	0.93	9.6923
18	0.1168	0.96	9.7496
19	0.1168	0.98	9.8002
20	0.1168	0.99	9.8449
21	0.1168	0.99	9.8844
22	0.1168	1	9.9193
23	0.1168	1	9.9501
24	0.1168	1	9.9774
25	0.1168	1	10.0014
26	0.1168	1	10.0227
27	0.1168	1	10.0414
28	0.1168	1	10.0580
29	0.1168	1	10.0727
30	0.1168	1	10.0856
31	0.1168	1	10.0970

Table S4. List of parameter values in the sandbar shark operating model. Steepness (h) and von Bertalanffy growth parameters were taken from SEDAR (2017).

Parameter	Value	
Survival-based stock–recruitment		
Z_0	1.0715	
Z_{min}	0.1604	
S_{max}	0.8518	
S_0	0.3377	
Z_{frac}	0.8503	
β	0.3658	
h	0.3	
$log(R_0)$	6.9078	
R_0	1000	
$Npups_0$	2919.848	
von Bertalanffy growth		
	Male	**Female**
K_{avg}	0.15	0.12
$L_{\infty\ avg}$	172.97	181.15
t_0	–2.33	–3.09

Table S5. Table describing how selectivity was modeled in each EM. Selectivity parameters that were fixed are denoted by values in the "Fixed params" columns. Shaded boxes indicate selectivity patterns with fewer parameters. Note that Atlantic sharpnose shark survey selectivities were estimated, but with tighter priors (i.e., lower standard deviations).

Atlantic sharpnose shark			**Fixed params - Female**						**Fixed params - Male offset**					
EM	**Survey selectivity**	**Time blocks?**	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**
all EMs - 1 fishery	F_1 - double normal	No												
CI EM - 3 surveys	S_1 - logistic	No												
	S_2 - logistic													
	S_3 - logistic													
DFA EM - 3 surveys	S_{DFA} - logistic	No												
CI EM - 4 surveys	S_1 - logistic	No												
	S_2 - logistic													
	S_3 - logistic													
	S_4 - logistic													
DFA EM - 4 surveys	S_{DFA} - logistic	Yes												
Sandbar shark			**Fixed params - Female**						**Fixed params - Male offset**					
EM	**Survey Selectivity**	**Time blocks?**	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**
all EMs - 4 fisheries	F_1 - double normal	No		–10			–999			2.95				1
	F_2 - logistic	No												
	F_3 - double normal	No		–10		1.93	–999						–999	1
	F_4 - logistic	No	–5	1										
CI EM - complete data	S_1 - double normal	No												1
	S_2 - logistic													
	S_3 - double normal			–10			–10							1
	S_4 - double normal			–10										
	S_5 - double normal			–10			–10	–999	–1					
	S_6 - double normal			–10			–999	–999				3.4	–1.9	
	S_7 - double normal			–10			–999							
DFA EM - complete data	S_{DFA} - random walk	No												
CI EM - missing data	S_1 - double normal	No												1
	S2 - logistic													
	S_3 - double normal			–10			–10							1
	S_4 - double normal			–10										
	S_5 - double normal			–10			–10	–999	–1					
	S_6 - double normal			–10			–999	–999				3.4	–1.9	
	S_7 - double normal			–10			–999							
DFA EM - missing data	S_{DFA} - random walk	Yes												