Dynamic factor analysis to reconcile conflicting survey indices of abundance

Cassidy D. Peterson, Michael J. Wilberg, Enric Corte´s, and Robert J. Latour

SEDAR77-RD53

Received: 5/26/2022

ICES Journal of Marine Science

ICES Journal of Marine Science (2021), 78(5), 1711–1729. doi:10.1093/icesjms/fsab051

Dynamic factor analysis to reconcile conflicting survey indices of abundance

Cassidy D. Peterson [1,*†], Michael J. Wilberg [2], Enric Cortés[3], and Robert J. Latour[1]

[1]*Virginia Institute of Marine Science, William & Mary, PO Box 1346, Gloucester Point, VA 23062, USA*
[2]*Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, PO Box 38, Solomons, MD 20688, 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.
[†]*Present address: Southeast Fisheries Science Center, National Marine Fisheries Service, 101 Pivers Island Road, Beaufort, NC 28516, USA.*

Peterson, C. D., Wilberg, M. J., Cortés, E., and Latour, R. J. Dynamic factor analysis to reconcile conflicting survey indices of abundance. – ICES Journal of Marine Science, 78: 1711–1729.

Received 31 August 2020; revised 2 March 2021; accepted 3 March 2021; advance access publication 2 May 2021.

Stock-wide trends in fish relative abundance are challenging to obtain when a single, comprehensive survey is unavailable, and multiple, spatially, and/or temporally fragmented surveys must be relied upon instead. Indices of abundance from multiple surveys frequently have differing trends, resulting in obscured true abundance patterns of the resource. We use an age-structured simulation model of two coastal shark species in the southeast United States to explore the performance of dynamic factor analysis (DFA) for reconciling multiple indices of abundance that are in conflict. Survey-specific time-variation in catchability was induced to generate conflicting indices of abundance. Key simulation sensitivities included survey variability, abundance pattern in the resource, and missing years of survey data. We caution against using DFA when there is no contrast in the underlying stock abundance or when trends in catchability in all surveys result in no survey that is representative of stock abundance. When multiple representative surveys were available, DFA proved useful across species in estimating stock-wide trends from conflicting survey indices with different selectivities, catchabilities, variances, and, to a lesser extent, with missing data. Our results suggest that resolving contrasting patterns among multiple time-series of relative abundance can improve understanding of the temporal trend in stock abundance.

Keywords: catchability, coastal shark, data conflict, fishery-independent sampling, relative abundance

Introduction

Fishery-independent trends in relative abundance generated from scientific surveys are one of the most important inputs to a stock assessment, as they theoretically represent changes in stock size through the assumption that the resulting indices are proportional to total abundance (Hilborn and Walters, 1992; Francis, 2011). However, for spatially wide-ranging species or species that cross domestic and international management boundaries, comprehensive population-wide surveys are rarely available. As a result, fisheries scientists frequently rely on data collected from several independent and spatially fragmented survey programmes to estimate the patterns of abundance of migratory or transboundary species. Operationally, multiple survey indices of relative abundance are frequently input into a single stock assessment model, under the assumption that each survey index provides representative information about the underlying abundance of the stock (Conn, 2010a; Cortés, 2011; Cortés *et al.*, 2015).

Given the underlying complexities in fish habitat utilization and movement, including seasonal, ontogenetic, and sex-specific variation, each spatially restricted survey may not provide a representative signal of temporal trends in stock size (Conn, 2010a; Francis, 2011). Relatively spatially narrow surveys can indicate trends in local relative abundance that are not representative of the total stock (Maunder *et al.*, 2006; Wilberg *et al.*, 2009), and inherently low encounter rates can lead to difficulties measuring relative abundance (Cook, 2010). Surveys that encounter target

species along migratory routes must ensure a spatio-temporal match between sampling and availability of the resource to ensure adequate characterization of relative abundance. Furthermore, surveys that only sample a particular portion of the life cycle (e.g. nurseries, migratory gravid females) or with a limited selectivity are unlikely to reflect trends in total abundance (Maunder *et al.*, 2006; Conn, 2010a).

Despite sampling the same population, relative abundance trends derived from small-scale surveys frequently conflict, obscuring identification of the true trend in stock abundance and hindering interpretation and assessment performance (Francis, 2011; Cortés *et al.*, 2015; Maunder and Piner, 2017). In the context of conflicting survey indices, it is likely that one or more indices are not representative of the underlying pattern in stock abundance and should not be considered with the same weight as more representative indices (Schnute and Hilborn, 1993; Conn, 2010b). Hence, a more reliable way to assess and interpret wide-scale abundance of highly migratory and transboundary species is needed (Conn, 2010a).

Coastal sharks along the southeast coast of the United States are an example of migratory species that are challenging to monitor and assess due to their large home ranges and complex migratory and habitat utilization patterns (Pilling *et al.*, 2008; Cortés, 2011). Given their low economic value, resources supporting broad-scale shark surveys are not prioritized and are sometimes logistically infeasible (Stevens *et al.*, 2000; Field *et al.*, 2009; Cortés *et al.*, 2015). Consequently, many shark surveys in the United States are constrained within state boundaries, resulting in survey index trends that reflect measures of localized relative abundance across spatial ranges much smaller than the actual distributions of the target population. For example, in the southeastern United States, shark-directed surveys sample only a few fixed stations or exclusively sample nursery areas, necessitating additional reliance on non-directed surveys of various gear types to supplement abundance information. The spatial sparseness of these data does not allow detailed spatial analyses for the full range of most stocks. Including many disagreeing survey indices that are each representative of local abundance in a stock assessment model violates the assumption that indices of abundance are proportional to total stock abundance. Relative abundance from shark surveys also typically includes impossibly large inter-annual variability given the shark's life history strategy (Cortés *et al.*, 2015). Time-series of shark relative abundance often conflict, likely due to inherent noise (annual variability and/or observation error; Cortés *et al.*, 2015) and complex life cycles (i.e. multi-year reproductive periodicity and accompanying movement patterns, ontogenetic and sex-specific habitat utilization, etc.; Ellis *et al.*, 2008; Grubbs, 2010; Conrath and Musick, 2012; Simpfendorfer and Heupel, 2012), compounded by a variable spatiotemporal mismatch between annual movement and sampling (Conn, 2010a). Due to lagging life history, movement, and comprehensive abundance data, coastal sharks are routinely considered data-limited species.

Dynamic factor analysis (DFA) is a multivariate dimension reduction technique designed to detect common, latent trends from a collection of time-series. The approach can accommodate data frequently encountered in ecology, including short, non-stationary, covarying time-series with missing years of data (Zuur *et al.*, 2003a; Holmes *et al.*, 2014). The application of DFA has increased in the marine ecological and fisheries literature (e.g. Colton *et al.*, 2014; Stachura *et al.*, 2014; Buchheister *et al.*, 2016; Jorgensen *et al.*, 2016; Ouellet *et al.*, 2016; Latour *et al.*, 2017), and DFA has frequently been used to analyse multiple noisy indices of relative abundance (e.g. Zuur *et al.*, 2003b; Zuur and Pierce, 2004; Chen *et al.*, 2006; Azevedo *et al.*, 2008; Peterson *et al.*, 2017a).

Through simulation analyses, we tested DFA as a method to reconcile conflicting indices of abundance and thereby clarify abundance trends of a stock under a variety of scenarios using two representative coastal sharks off the southeast United States: the large coastal sandbar shark (*Carcharhinus plumbeus*) and the small coastal Atlantic sharpnose shark (*Rhizoprionodon terraenovae*). These two species were selected because they are the most data-rich species of their respective management units. Survey indices were simulated to be in conflict by inducing time-varying patterns in the catchability coefficient (q or the proportionality constant between survey catch and abundance), thereby violating the implicit assumption of constant proportionality in survey programmes. The applicability of DFA to reconcile survey indices was also simulated under variable fishing mortality patterns, numbers of surveys, survey index variability, and with missing years of survey data. Based on our simulations, we present recommendations for using DFA as a method of reconciling conflicting survey indices.

Methods

Modelled species

Female Atlantic sharpnose sharks reach median sexual maturity at 1.6 years (Loefer and Sedberry, 2003), have a maximum longevity of 23 years (Frazier *et al.*, 2014), reproduce annually, and have an average of 4–5 pups per litter (Castro, 2009). Atlantic sharpnose sharks off the US Atlantic coast and within the Gulf of Mexico were most recently assessed as a single stock in SouthEast Data, Assessment, and Review (SEDAR) 34 (SEDAR, 2013) using a single-sex, state-space, age-structured production model. Genetic evaluation has since revealed Atlantic sharpnose shark stock structure between the Gulf of Mexico and Atlantic Ocean (Davis *et al.*, 2019). Single-sex life history parameters used in our simulation were based on individuals from the Atlantic Ocean (SEDAR, 2013; Supplementary Tables S1 and S2).

Female sandbar sharks reach sexual maturity at a median age of 14 years (Baremore and Hale, 2012), longevity is estimated to be 31 years (SEDAR, 2017), average fecundity is eight pups (Baremore and Hale, 2012), and the reproductive cycle has been proposed to be two or three years (Baremore and Hale, 2012) and is, therefore, modelled as 2.5 years (SEDAR, 2011). Sandbar sharks comprise a single genetic stock in the Atlantic Ocean and within the Gulf of Mexico (Heist *et al.*, 1995). The most recent assessment of the sandbar shark was conducted using Stock Synthesis (SEDAR, 2017), representing a pseudo-update from SEDAR 21 (SEDAR, 2011). Sex-specific life history parameters used for our simulations were obtained from SEDAR 54 (SEDAR, 2017; Supplementary Tables S3 and S4).

Operating model

Two separate operating models (OMs) were developed, one for each species of interest. Due to a lack of sex-specific information for the Atlantic sharpnose shark, a one-sex OM was created for this species to reflect our limited understanding of sex-specific

dynamics within the stock. A single fishery was simulated where the selectivity was defined as the average selectivity over the four fisheries from SEDAR (2013), and either three or four surveys were implemented. The sandbar shark OM was a two-sex model, with four fisheries and seven surveys (SEDAR, 2017), due to the availability of additional biological and fishery information.

The OM for both species was developed using an age-structured model:

$$N_{s,a,y+1} = \begin{cases} R_{s,y+1} & a = 1 \\ N_{s,a-1,y}\mathrm{e}^{-Z_{s,a,y}} & 1 < a < A, \\ \left(N_{s,A-1,y}\mathrm{e}^{-Z_{s,A-1,y}}\right) + \left(N_{s,A,y}\mathrm{e}^{-Z_{s,A,y}}\right) & a = A \end{cases} \tag{1}$$

where $N_{s,a,y}$ represents abundance of individuals of sex s at age a and year y, $R_{s,y}$ is recruitment, $Z_{s,a,y}$ is sex-, age-, and year-specific total mortality, the age at recruitment is 1, and A is the age of the plus group ($A = 18$ for Atlantic sharpnose shark, $A = 31$ for sandbar shark). In these simulated populations, fishing mortality occurs simultaneously with natural mortality, and catch is equal to landings, which functionally assumes no discarding.

Sex-specific annual recruitment ($R_{s,y}$) was calculated as the number of pups (*Npups*) produced adjusted by the annual age-0 survival:

$$R_{s,y+1} = Npups_y \times e^{-M_{0,\ y}}. \tag{2}$$

In the two-sex sandbar shark model, we assumed a 1:1 male: female ratio at birth (SEDAR, 2017) and that the age-0 survival rate was equal for males and females.

Density dependence was implemented through age-0 instantaneous natural mortality,

$$M_{0,y} \sim N\left\{\left[1 - \left(\frac{Npups_y}{Npups_0}\right)^{\beta} \times (Z_{min} - Z_0) + Z_0\right], \sigma_M\right\}, \tag{3}$$

following the low-fecundity, survival-based recruitment function (LFSR) of Taylor *et al.* (2013), with added process uncertainty (arbitrarily defined $\sigma_M = 0.1$) and where $Npups_y$ is the yearly spawning output (in number of embryos), $Npups_0$ is the spawning output (number of embryos produced) at virgin conditions, β is a shaping parameter controlling the density-dependent survival of prerecruits (*Npups*), Z_{min} is the instantaneous total mortality that corresponds to maximum survival of prerecruits (*Npups*), and Z_0 is the instantaneous total mortality of prerecruits at virgin conditions. The number of pups produced each year can be determined by multiplying yearly abundance of mature females-at-age by their age-specific fecundity summed across all ages:

$$\begin{gathered} Npups_y = \sum_a N_{f,a,y} \times \frac{1}{RP} \times p_{f,a} \times f_a, \\ p_{f,a} \sim N(p_{avg_{f,a}},\ p_{avg_{f,a}} \times CV_p), \\ f_a \sim N(f_{avg_a},\ f_{avg_a} \times CV_f), \end{gathered} \tag{4}$$

where RP represents the reproductive periodicity (RP =1 for Atlantic sharpnose sharks and RP =2.5 for sandbar sharks), $p_{f,a}$ represents the proportion of mature females-at-age, and f_a represents the fecundity-at-age, with associated process error, while $p_{avg,f,a}$ and $f_{avg,a}$ are static average maturity-at-age and fecundity-at-age vectors. Additional uncertainty in maturity- and fecundity-at-age was based on arbitrarily defined coefficients of variation ($CV = \sigma/\mu$), where the added uncertainty was selected to generally remain equivalent to or smaller than the differences in maturity and fecundity between ages ($CV_p = 0.01$, $CV_f = 0.1$). (See Supplementary material for more information on parameterizing the spawner–recruit function.)

Sex-specific catch-at-age ($C_{s,a,y,i}$) in each year was calculated across all fleets (i):

$$C_{s,a,y,i} = N_{s,a,y} \times \frac{F_{s,a,y,i}}{Z_{s,a,y}} \times \left(1 - e^{-Z_{s,a,y}}\right), \tag{5}$$

where $F_{s,a,y,i}$ is instantaneous fishing mortality-at-sex, -age, -year, and -fleet and $Z_{s,a,y}$ represents sex-specific total instantaneous mortality-at-age and -year. Sex- and age-specific annual instantaneous fishing mortality for each fleet was calculated as:

$$F_{s,a,y,i} = sel_{s,a,i} \times F_y \times \rho_i, \tag{6}$$

where ρ_i represents proportion of fishing mortality attributed to fleet i. Fleet selectivity-at-sex and -age ($sel_{s,a,i}$) and ρ_i were constant over time.

Total sex-specific instantaneous mortality-at-age and -year was calculated by summing fishing mortality over fleets and natural instantaneous mortality-at-age:

$$\begin{gathered} Z_{s,a,y} = \sum_{\forall i} F_{s,a,y,i} + M_{s,a}, \\ M_{s,a} \sim N(M_{avg\,s,a},\ M_{avg\,s,a} \times 0.01), \end{gathered} \tag{7}$$

where sex-specific instantaneous natural mortality-at-age ($M_{s,a}$) was stochastically implemented based on the time-invariant, average mortality-at-age vector $M_{avg\_s,a}$.

Selectivity-at-age for the Atlantic sharpnose shark was obtained from SEDAR (2013). Fishery selectivity was dome-shaped, while selectivity for each survey followed a logistic curve. Sex-specific, length-based selectivity of sandbar sharks was available in SEDAR (2017). Length-based selectivity was converted to selectivity-at-age by converting length bins to age using sex-specific von Bertalanffy growth parameters, then fitting selectivity-at-age using Stock Synthesis selectivity helper Excel spreadsheets (available via https://vlab.ncep.noaa.gov/web/stock-synthesis/document-library, accessed June 2020). All sandbar shark fisheries and surveys followed double-normal or logistic selectivity, and the structural form of the selectivity curve was preserved through the conversion to selectivity-at-age. Indices of relative abundance for each sex, age, year, and survey ($I_{s,a,y,j}$) were assumed to be related to abundance via:

$$\begin{gathered} I_{s,a,y,j} = q_{y,j} v_{s,a,j} N_{s,a,y} \times \exp\left(\epsilon_{s,a,y,j} - \frac{\sigma^2_{s,a,y,j}}{2}\right), \\ \epsilon_{s,a,y,i} \sim N(0,\ \sigma_{s,a,y,j}), \\ \sigma_{s,a,y,j} = q_{y,j} v_{s,a,j} N_{s,a,y} CV_j, \\ CV_j \sim U(CV_{avg_j} - 0.1,\ CV_{avg_j} + 0.1), \end{gathered} \tag{8}$$

where $q_{y,j}$ represents survey-specific catchability coefficient in each year, $v_{s,a,j}$ is sex-, age-, and survey-specific selectivity-at-age that was assumed to be constant over time, and standard deviation, $\sigma_{s,a,y,j}$, is defined in terms of the survey-specific CV_j. Survey uncertainty was implemented with a CV_j for each survey j that varies uniformly around a survey-specific average CV ($CV_{avg\_j} \pm$

0.1). Sex-specific indices were summed across age to produce the final annual index for each survey:

$$I_{y,j} = \sum_{\forall a} I_{f,a,y,j} + I_{m,a,y,j}, \quad (9)$$

where subscripts f and m indicate female and male indices, respectively. Note that the Atlantic sharpnose shark scenario was a female-only model.

Models were provided an arbitrary starting population size of recruits (in relative units of number of individuals), propagated to numbers-at-age using age-specific M_a. The deterministic form of the simulation was run for 1000 years in the absence of fishing mortality, ensuring that the population was equilibrated with respect to the LFSR parameters at virgin conditions. This deterministically derived abundance-at-age vector was used as equilibrium conditions for all subsequent fishing scenarios.

Simulated trials

Including two species in our study allowed us to demonstrate how successfully DFA reconciled conflicting survey indices across life history strategies and with variable data availability. Within the Atlantic sharpnose shark simulation, we explored how DFA performance was affected by (1) underlying stock abundance through altered patterns in fishing mortality, (2) survey variability via altered index CV, (3) number of surveys by adding a partial, fourth survey, and (4) the presence of disagreeing surveys, generated by inducing knife-edged and gradual shifts in survey catchability (Table 1 and Figure 1).

The additional complexity of the sandbar shark OM enabled more intricate exploration of the effects of conflicting indices on the performance of DFA. Within the sandbar shark DFA simulation, we more comprehensively explored (4) the effects of changing survey catchability patterns, as well as (5) the effects of missing years of survey data. Variable combinations of the number of surveys that underwent changes in q (i.e. 0–3 surveys) and the directionality of the change in q (i.e. increasing q or decreasing q) were considered (Table 2 and Figure 1). Note that missing data were based on observed survey indices within SEDAR (2017) and included indices with missing years and surveys that only span a fraction of the total simulated time-span (Figure 1). To demonstrate how DFA performs when all survey data have trends in catchability (i.e. all surveys undergo shifts in survey catchability, with no surveys proportional to abundance), we simulated four additional (6) "all time-varying catchability" scenarios for the sandbar shark. In each "all time-varying catchability" scenario, five surveys had trends in survey catchability in one direction and the remaining two surveys trended in the opposite direction. "All time-varying catchability" scenarios were simulated with and without missing data (Table 1).

Simulations spanned 65 or 90 years for the Atlantic sharpnose and sandbar sharks, respectively. Each trial was simulated 100 times to account for variable combinations of uncertainty, and the number of simulations was kept relatively small due to the complexities of the DFA rescaling approach (detailed below) and to allow for more in-depth analysis of each trial. Subsequent analyses were based on simulated years 40–65 for the Atlantic sharpnose shark and 51–90 for the sandbar shark (Figure 1), allowing for a "burn in" period to enable equilibration to the fishing mortality scenarios (outlined in Table 1) consistent with the intrinsic population growth rate and representing a more realistic time-series duration for each species considered.

Atlantic sharpnose shark shifts in survey catchability were either knife-edged (happening in year 50) or gradual (linear ramp over 10–15 years starting in year 50), and survey catchability shifted from 0.00025 to 0.001 (q_1) or 0.003 to 0.004 (q_3). Sandbar shark patterns in survey catchability ranged from 0.01 to 0.045 or 0.045 to 0.01from years 51–90 in either a jittered linear or stepwise (change of 0.005 every five years) manner (Figure 1). In the "all time-varying catchability" sandbar shark scenarios, four survey catchability patterns experienced random noise superimposed on the linear trend for added variability.

Induced shifts and patterns in survey catchability were generally sufficiently large in magnitude to generate survey indices that displayed clearly observable contrast within our simulated scenarios. Though alternate models for generating time-varying survey catchability were considered, including those produced via random walks, they resulted in indices with greater agreement. Consequently, these scenarios were designed to include substantial directional changes that have been shown to cause bias in age-structured stock assessment models (Wilberg and Bence, 2006).

DFA

The general form of a DFA model can be written as follows, adopting the notation of Zuur *et al.* (2003a):

$$\begin{aligned} y_t &= \Gamma\alpha_t + \varepsilon_t, \quad \text{where } \varepsilon_t \sim \text{MVN}(0, \boldsymbol{H}), \\ \alpha_t &= \alpha_{t-1} + \eta_t, \quad \text{where } \eta_t \sim MVN(0, \boldsymbol{Q}), \end{aligned} \quad (10)$$

where y_t is a vector of n standardized (Z-scored) time-series at each time t, and α_t is a vector of m common trends (where $m < n$) that are modelled as a random walk with associated error (η_t; Zuur *et al.*, 2003a, b; Holmes *et al.*, 2014), $\boldsymbol{\Gamma}$ is an $n \times m$ matrix containing factor loadings, which indicate how strongly each resulting trend influences each time-series (Zuur *et al.*, 2003a, b; Holmes *et al.*, 2014). Factor loadings greater than or equal to $|0.2|$ indicate that the resulting DFA trend strongly describes the corresponding input time-series (Zuur *et al.*, 2003b; Tam *et al.*, 2013). Both observation and process error terms, ε_t and η_t, assume a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrices $\boldsymbol{H}$ and $\boldsymbol{Q}$, respectively. To ensure that the model is identifiable, $\boldsymbol{Q}$ is set equal to the identity matrix $\boldsymbol{I}$, while $\boldsymbol{H}$ is free to take on various forms (Holmes *et al.*, 2014). All factor loadings, common trends, and fitted values are unitless (Zuur *et al.*, 2003b).

In the current study, diagonal elements of the $\boldsymbol{H}$ matrix (observation variances) were defined as the average survey CV, and the off-diagonal elements were set to zero, functionally assuming that each time-series was independent (no covariance between survey indices). We elected to use survey CV in place of survey variances in the $\boldsymbol{H}$ matrix, because CVs are adjusted for the magnitude of the index; furthermore, in practice, using survey variances in the $\boldsymbol{H}$ matrix can result in overfitting or inflated uncertainty (C. Peterson, unpublished data). Though this was a simplifying assumption made in the current analyses, alternative assumptions could be made in practice, including allowing the model to estimate observation variance and covariance (e.g. Peterson *et al.*, 2017a). By defining the elements of the $\boldsymbol{H}$ matrix, we are

Table 1. List of trials simulated for the Atlantic sharpnose shark.

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
ConstF 3S	**ConstF_const_1**	Const F ($F = 0.2$)	0.5	0.5	0.5	const q	const q	const q	0.5	const q	**ConstF_const_1_4**	ConstF 4S
	ConstF_const_2	Const F ($F = 0.2$)	0.3	0.5	0.7	const q	const q	const q	0.5	const q	**ConstF_const_2_4**	
	ConstF_const_3	Const F ($F = 0.2$)	0.7	0.5	0.3	const q	const q	const q	0.5	const q	**ConstF_const_3_4**	
ConstF 3S-Δq	**ConstF_knife_1**	Const F ($F = 0.2$)	0.5	0.5	0.5	knife ↑q	const q	knife ↓q	0.5	const q	**ConstF_knife_1_4**	ConstF 4S-Δq
	ConstF_knife_2	Const F ($F = 0.2$)	0.3	0.5	0.7	knife ↑q	const q	knife ↓q	0.5	const q	**ConstF_knife_2_4**	
	ConstF_knife_3	Const F ($F = 0.2$)	0.7	0.5	0.3	knife ↑q	const q	knife ↓q	0.5	const q	**ConstF_knife_3_4**	
	ConstF_grad_1	Const F ($F = 0.2$)	0.5	0.5	0.5	grad ↑q	const q	grad ↓q	0.5	const q	**ConstF_grad_1_4**	
	ConstF_grad_2	Const F ($F = 0.2$)	0.3	0.5	0.7	grad ↑q	const q	grad ↓q	0.5	const q	**ConstF_grad_2_4**	
	ConstF_grad_3	Const F ($F = 0.2$)	0.7	0.5	0.3	grad ↑q	const q	grad ↓q	0.5	const q	**ConstF_grad_3_4**	
IncF 3S	**IncF_const_1**	↑F ($F = 0/0.4$)	0.5	0.5	0.5	const q	const q	const q	0.5	const q	**IncF_const_1_4**	IncF 4S
	IncF_const_2	↑F ($F = 0/0.4$)	0.3	0.5	0.7	const q	const q	const q	0.5	const q	**IncF_const_2_4**	
	IncF_const_3	↑F ($F = 0/0.4$)	0.7	0.5	0.3	const q	const q	const q	0.5	const q	**IncF_const_3_4**	
IncF 3S-Δq	**IncF_knife_1**	↑F ($F = 0/0.4$)	0.5	0.5	0.5	knife ↑q	const q	knife ↓q	0.5	const q	**IncF_knife_1_4**	IncF 4S-Δq
	IncF_knife_2	↑F ($F = 0/0.4$)	0.3	0.5	0.7	knife ↑q	const q	knife ↓q	0.5	const q	**IncF_knife_2_4**	
	IncF_knife_3	↑F ($F = 0/0.4$)	0.7	0.5	0.3	knife ↑q	const q	knife ↓q	0.5	const q	**IncF_knife_3_4**	
	IncF_grad_1	↑F ($F = 0/0.4$)	0.5	0.5	0.5	grad ↑q	const q	grad ↓q	0.5	const q	**IncF_grad_1_4**	
	IncF_grad_2	↑F ($F = 0/0.4$)	0.3	0.5	0.7	grad ↑q	const q	grad ↓q	0.5	const q	**IncF_grad_2_4**	
	IncF_grad_3	↑F ($F = 0/0.4$)	0.7	0.5	0.3	grad ↑q	const q	grad ↓q	0.5	const q	**IncF_grad_3_4**	
DecF 3S	**DecF_const_1**	↓F ($F = 0.4/0$)	0.5	0.5	0.5	const q	const q	const q	0.5	const q	**DecF_const_1_4**	DecF 4S
	DecF_const_2	↓F ($F = 0.4/0$)	0.3	0.5	0.7	const q	const q	const q	0.5	const q	**DecF_const_2_4**	
	DecF_const_3	↓F ($F = 0.4/0$)	0.7	0.5	0.3	const q	const q	const q	0.5	const q	**DecF_const_3_4**	
DecF 3S-Δq	**DecF_knife_1**	↓F ($F = 0.4/0$)	0.5	0.5	0.5	knife ↑q	const q	knife ↓q	0.5	const q	**DecF_knife_1_4**	Dec 4S-Δq
	DecF_knife_2	↓F ($F = 0.4/0$)	0.3	0.5	0.7	knife ↑q	const q	knife ↓q	0.5	const q	**DecF_knife_2_4**	
	DecF_knife_3	↓F ($F = 0.4/0$)	0.7	0.5	0.3	knife ↑q	const q	knife ↓q	0.5	const q	**DecF_knife_3_4**	
	DecF_grad_1	↓F ($F = 0.4/0$)	0.5	0.5	0.5	grad ↑q	const q	grad ↓q	0.5	const q	**DecF_grad_1_4**	
	DecF_grad_2	↓F ($F = 0.4/0$)	0.3	0.5	0.7	grad ↑q	const q	grad ↓q	0.5	const q	**DecF_grad_2_4**	
	DecF_grad_3	↓F ($F = 0.4/0$)	0.7	0.5	0.3	grad ↑q	const q	grad ↓q	0.5	const q	**DecF_grad_3_4**	
UF 3S	**UF_const_1**	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_const_1_4**	UF 4S
	UF_const_2	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_const_2_4**	
	UF_const_3	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_const_3_4**	
UF 3S-Δq	**UF_knife_1**	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_knife_1_4**	UF 4S-Δq
	UF_knife_2	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_knife_2_4**	
	UF_knife_3	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_knife_3_4**	
	UF_grad_1	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_grad_1_4**	
	UF_grad_2	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_grad_2_4**	
	UF_grad_3	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_grad_3_4**	

F is instantaneous fishing mortality pattern, I represents index of relative abundance indexed by survey (S) number, CV is coefficient of variation, and q is catchability coefficient. The trial name is given by three values: F [constant, increasing, decreasing, or U-shaped (increasing then decreasing)], q (constant, knife-edged, or gradual), and survey number (1, 2, or 3). ↑ represents increasing patterns in q or F while ↓ represents decreasing patterns in q or F.

[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, fourth index started at year 55.

accounting for known uncertainty as calculated from survey index standardization. Explanatory variables were not considered. While the number of common trends produced by the model can vary (where $m < n$), we defined $m = 1$, because we were only interested in obtaining a single trend representative of the underlying trend in the population.

The fundamental assumptions of a DFA model are equivalent to those of a linear regression, thereby including normality, independent error, and homogeneity of residuals (Zuur *et al.*, 2003b). Models were fitted using the state-space multivariate autoregressive modelling package "MARSS" in R (Holmes *et al.*, 2018).

Rescaling approach

It is generally recommended that time-series be detrended and standardized (z-scored) prior to conducting a DFA, which often includes log-transformation of time-series with lognormal error (e.g. survey indices; Zuur *et al.*, 2003a, b; Holmes *et al.*, 2014). Subsequently, the resulting DFA trend, which is often the solution we are most interested in for stock assessment purposes, is generated in log-space, on a unitless scale that spans positive and negative values. Under this conventional approach to data manipulation in preparation for DFA, we are unable to back-transform resulting DFA trends out of an abstract, log- and unitless-space for successive analyses and interpretation.

We developed a unique approach to rescale the input survey indices in a manner consistent with the requirements of DFA (z-score), while simultaneously preserving the error structure and the relative scale of the survey indices, and allowing backtransformation of the resulting DFA index out of detrended, log-scale. The rescaling approach is as follows: (1) each survey index (I_j) was multiplied by a constant (c_j, see following paragraph), (2) all

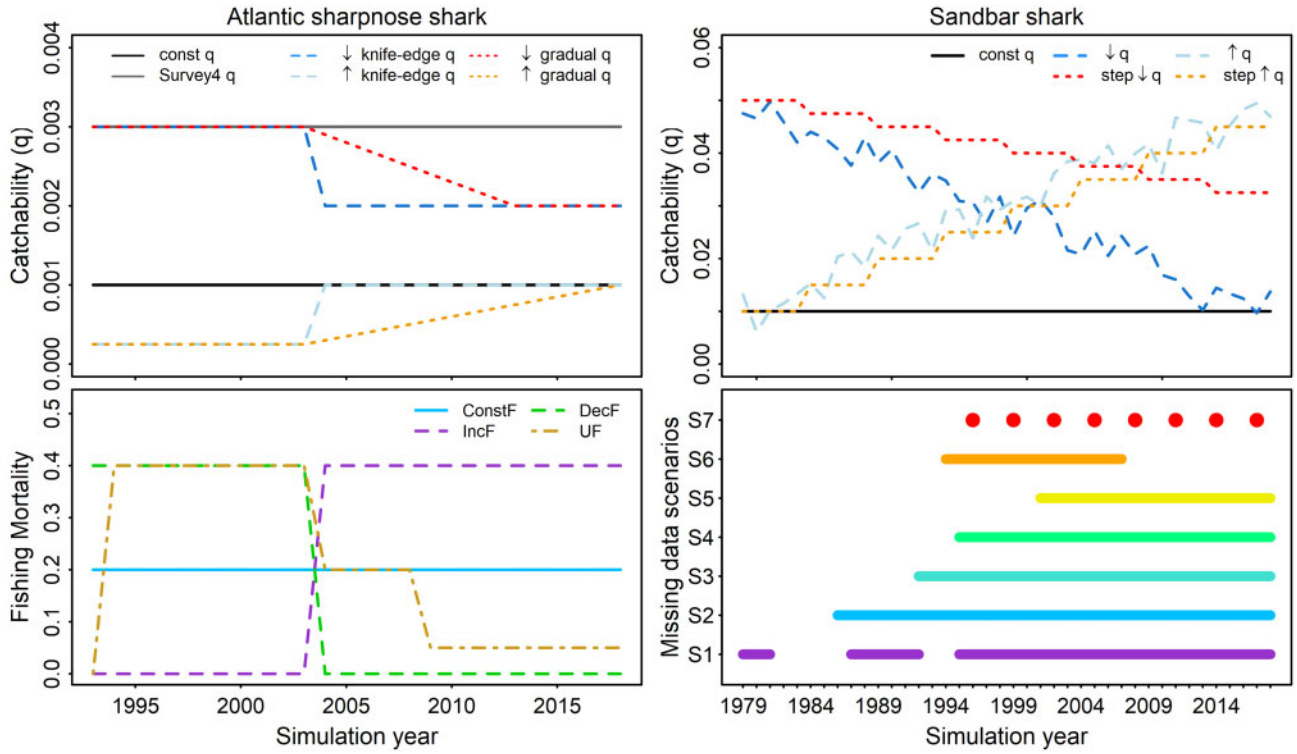

Figure 1. Alternate simulation scenarios for the Atlantic sharpnose shark (left) and sandbar shark (right) including various time-varying survey catchability configurations (top row), fishing mortality patterns for Atlantic sharpnose shark simulations (bottom left), and available years of survey data in the "Missing data" scenario for the sandbar shark simulation for each survey (S; bottom right).

indices from step 1 were log-transformed, thereby normalizing survey error, (3) each log-transformed survey index from step 2 was centred 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 model was fitted, and (7) the resulting DFA-predicted trend was back-transformed by first multiplying by the *GSD* and then exponentiating with bias correction $\left\{\exp\left[GSD \times \alpha_t + \left(\frac{GSD \times SE_t^2}{2}\right)\right]\right\}$. 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 log-normal error of the DFA trend.

Prior to log-transforming, each index was multiplied by a constant c_j. The choice of c_j 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 $j = 1$ to n, and adjusting each c_j until the following conditions (a–d) were met within the steps of the rescaling approach outlined above:

(a) *The mean of each log-transformed survey index was >0.* Note that if the logged survey-specific mean is close to 0, then the rescaling approach would not work because we would essentially be dividing by 0 in step 3. Furthermore, 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_js that resulted in log-transformed survey index means $\gtrsim$2.5, when possible, produced the best results.

(b) *The resulting GSD was small (GSD≪1).* When the GSD is >1, 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 ≅1.* This ensures that the format of the input survey index most appropriately approximates a *z*-scored index, with a mean of 0 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 >1, 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 meet the above requirements, particularly when there are few input time-series, and unfortunately, the choice of $\boldsymbol{c}$ impacts the resulting DFA trend. The effect of the parameterization of $\boldsymbol{c}$ is greatest in one-way trips, whereas more complex patterns in underlying abundance are less affected. As such, constants should be chosen to 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. Maintenance of a

Table 2. Trials simulated for the sandbar shark simulation.

Figure 6 plotting labels	Figure 7 plotting labels	Trial	Missing data?	F[a]	q pattern	Trial	Missing data?	F[a]	q pattern
No change	No change	**SB1**	Yes	SB_F	const	**SB101**	No	SB_F	const
Change 1 q	I	**SB2**	Yes	SB_F	↑q S1	**SB102**	No	SB_F	↑q S1
		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
	D	**SB9**	Yes	SB_F	↓q S1	**SB109**	No	SB_F	↓q S1
		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
Change 2 q	I-I	**SB16**	Yes	SB_F	↑q S1-S3	**SB116**	No	SB_F	↑q S1-S3
		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
	D-D	**SB19**	Yes	SB_F	↓q S1-S3	**SB119**	No	SB_F	↓q S1-S3
		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
	I-D	**SB22**	Yes	SB_F	↑q S1,↓q S3	**SB122**	No	SB_F	↑q S1,↓q S3
		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
	D-I	**SB25**	Yes	SB_F	↓q S1,↑q S3	**SB125**	No	SB_F	↓q S1,↑q S3
		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
Change 3 q	I-I-I	**SB28**	Yes	SB_F	↑q S1-S3-S5	**SB128**	No	SB_F	↑q S1-S3-S5
		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
	D-D-D	**SB31**	Yes	SB_F	↓q S1-S3-S5	**SB131**	No	SB_F	↓q S1-S3-S5
		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
	I-D-I	**SB34**	Yes	SB_F	↑q S1, ↓q S3-S5	**SB134**	No	SB_F	↑q S1, ↓q S3-S5
		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
	D-I-D	**SB37**	Yes	SB_F	↓q S1, ↑q S3-S5	**SB137**	No	SB_F	↓q S1, ↑q S3-S5
		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
Change 7 q	5I-2D	**SB40**	Yes	SB_F	↑q S1-S3-S5-S6-S7, ↓ q S2-S4	**SB140**	No	SB_F	↑q S1-S3-S5-S6-S7, ↓q S2-S4
	5D-2I	**SB41**	Yes	SB_F	↓q S1-S3-S5-S6-S7, ↑q S2-S4	**SB141**	No	SB_F	↓q S1-S3-S5-S6-S7, ↑q S2-S4

"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. I or ↑ represents increasing patterns in q, while D or ↓ represents decreasing patterns in q. $CV_1 = 0.38$, $CV_2 = 0.48$, $CV_3 = 0.65$, $CV_4 = 0.24$, $CV_5 = 0.30$, $CV_6 = 0.36$, and $CV_7 = 0.40$.
[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, and 0.05 in years 76–100.

consistent and reliable trend that follows the raw survey indices is important. This is obviously challenging if the analyst is unsure which survey inputs are reliable, and the analyst may have to rely on best judgement (e.g. compare percent change in the backtransformed DFA predicted trend to the median percent change in all input survey indices). Note that this recommendation is not important if the magnitude of change in the resulting DFA index is not of interest.

The general pattern of each survey index should remain similar across each step of the rescaling protocol, and resulting DFA trends should look comparable to those generated from a traditionally scaled (survey indices are log-transformed then z-scored) DFA model. We also recommend ensuring that the mean fit ratio is low (where fit ratio$_j = \Sigma_t {y_{jt}}^2 / \Sigma_t {\varepsilon_{jt}}^2$ for each survey index j and the mean fit ratio is the average fit ratio across all surveys; see Supplementary material), as the mean fit ratio will appreciably increase with a less ideal vector of constants $\boldsymbol{c}$.

We emphasize that this rescaling approach is merely a substitute for z-scoring indices prior to implementation of the DFA. This approach preserves the features of a z-scored time-series (i.e. mean of 0 and standard deviation equal to 1) and ensures that the data transformation is equivalent for each input survey index such that the resulting DFA trend can be backtransformed. This rescaling process does not impact the DFA model itself but is instead an alternative form of data preparation.

Note that given the time-consuming, iterative process of identifying a vector of constants ($\boldsymbol{c}$) by which to multiply each survey index in our rescaling approach, we could not estimate a unique $\boldsymbol{c}$ for each iteration of every trial. Consequently, we estimated $\boldsymbol{c}$ for the first iteration of each scenario, and assumed that as a single vector for the entire scenario. Based on the simulation results of each iteration assuming the fixed $\boldsymbol{c}$, we went back to modify conspicuously incorrect vectors of $\boldsymbol{c}$s, as observed by obviously overfitted DFA trends (e.g. very jagged DFA trends that closely track the interannual variability in one survey but are poorly fitted to others) and very high SDs from step (*c*) above.

Analysing results

Comparison between simulated abundance trends and DFA-estimated trends required transformation due to the different scales of each trend. Consequently, we compared standardized back-transformed DFA-predicted trends to standardized total abundance using annual relative error $\left[relative\ error_y = \widehat{N_{std,y}} - N_{std,y},\ \forall y,\ \text{where } N_{std,y} = \frac{N_y - \overline{N_y}}{SD(N_y)}\right]$ and root-mean-squared-error (RMSE) for each trial (e.g. Conn, 2010a, b). Because each trend was standardized, annual relative error is centred on 0, and thus a wider spread in annual relative error is indicative of a more poorly fitted model. RMSE was calculated across years and each value was representative of a single simulation iteration. All analyses were conducted in R (version 3.6.2; R Core Team, 2019).

Results

We present the effects of (1) underlying pattern of population abundance, (2) survey variability, (3) the number of surveys, and (4) the presence of conflicting survey indices on the ability of DFA to reconcile disagreeing survey indices of relative abundance and thereby approximate population trends within the Atlantic sharpnose shark (SN) simulation (Table 1). The sandbar shark (SB) simulation allowed us to expand on the effect of the number of survey indices (3) and the presence of conflicting indices (4), as well as explore the effects of (5) missing data on DFA's ability to reconcile conflicting indices of relative abundance (Table 2). We also present DFA results from (6) the "all time-varying catchability" scenarios where all surveys experience survey catchability trends over time. Note that throughout the results section, when referring to DFA performance, we are referring to the ability of the DFA model to predict a resulting common trend that follows the same trend as the underlying population as measured by annual relative error and RMSE of the DFA-predicted trend compared to the simulated trend in abundance. Overall, DFA was generally able to successfully reconcile multiple survey indices, even when those indices were in conflict, and DFA performance was not affected by species' life history strategy (Figures 2 and 3).

Underlying pattern of population abundance

DFA performance was most affected by the underlying trend in population abundance in the Atlantic sharpnose shark simulation (Figure 4). When there was little contrast in population abundance (ConstF), DFA did not perform well, as demonstrated by high spread in annual relative error and high RMSE. When there was a one-way trip in the simulated population (i.e. IncF and DecF), DFA performed extremely well even across variable survey CVs, shifts in the survey catchability coefficient (q), and the number of surveys. When the population pattern was more complex (UF), DFA performance was good overall. However, performance declined in trials where surveys were simulated under certain combinations of survey uncertainty and shifts in the catchability coefficient (see next section; Figure 5 and Supplementary Figures S1 and S2).

Survey variability

The variability of each survey index was manipulated by altering survey-specific CVs in alternate Atlantic sharpnose shark simulation scenarios. Because we specified observation error within the DFA model (i.e. propagating known observation error), reduced survey CV functionally increased the weight of that survey on the resulting DFA trend, as observed by greater factor loadings. Overall, the effect of survey variability had relatively little impact on DFA performance on its own. However, the effects of survey CV became problematic when combined with shifts in survey catchability (Figure 5 and Supplementary Figures S1 and S2). For example, under a fishing mortality scenario in which there was little underlying contrast in the data (ConstF), when a given survey underwent a shift in survey catchability and the CV of that survey was smaller, the resulting DFA trend had a tendency to more closely follow the shifted trend in the survey as opposed to the true underlying constant population abundance trend. This tendency was also observed in more complex fishing mortality scenarios (UF_knife_2, UF_grad_2; SB129, SB135; Figure 5 and Supplementary Figures S1–S4) and exacerbated in examples with fewer surveys, resulting in increased annual relative error and RMSE (Figure 4 and Supplementary Figures S1 and S2).

Number of surveys

The addition of surveys improved DFA performance. Though the addition of an incomplete, fourth survey (with constant catchability) in the Atlantic sharpnose shark simulation did not have a particularly noticeable effect under many fishing mortality scenarios (i.e. ConstF and DecF), it improved DFA performance under the more complex fishing mortality scenario (UF where two surveys conflicted due to shifts in catchability coefficients; Figure 4 and Supplementary Figures S1 and S2).

Under similar population abundance trends and survey specific shifts in survey catchability, the spread in annual relative error and RMSE was slightly lower for the sandbar shark simulation compared to the Atlantic sharpnose shark simulation. Despite differences in the complexities of each simulation, this improvement in DFA performance was likely due to the greater number of surveys in the sandbar shark simulation (seven) compared to the Atlantic sharpnose shark simulation (three or four; Figure 6).

Presence of conflicting survey indices

Conflicting survey indices were generated by introducing time-varying patterns in the catchability coefficient q for selected surveys in both the Atlantic sharpnose and sandbar shark simulations (Tables 1 and 2). When all survey indices agreed (i.e. no change in survey catchability), DFA performed very well, though performance declined in the absence of contrast in the underlying stock abundance.

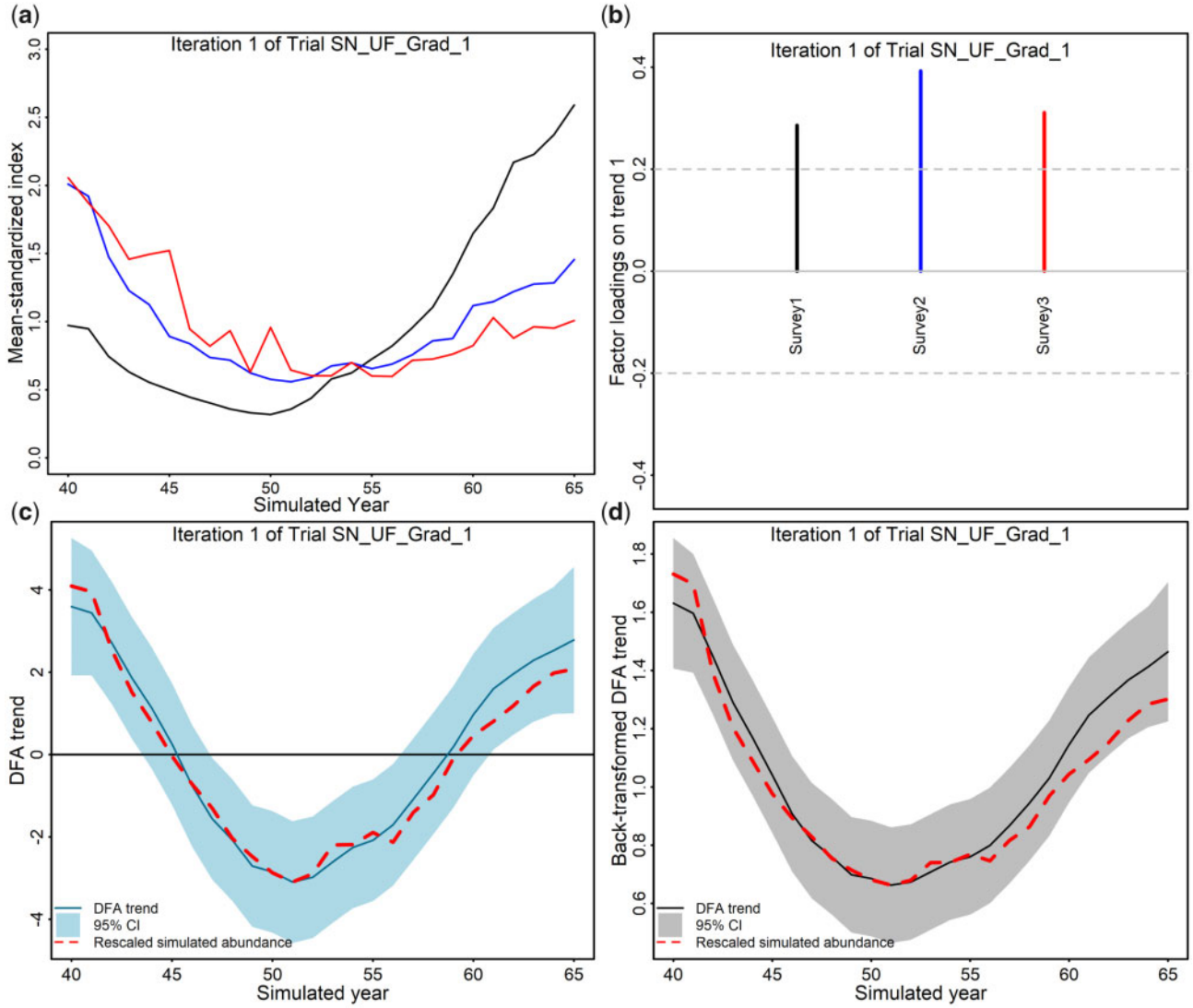

Figure 2. Example DFA model run for the first iteration of Atlantic sharpnose shark Trial UF_grad_1, including (a) input data as shown as mean-standardized survey indices, (b) corresponding factor loadings, denoting the strength of influence of the resulting DFA-predicted trend on each survey, (c) resulting 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 three input survey indices, two undergo shifts in catchability in opposing directions (black and red indices), while one survey experiences constant catchability (blue index).

In scenarios where DFA performed well in the absence of conflicting survey indices, overall performance remained relatively strong when conflicting indices were introduced. For example, in the Atlantic sharpnose shark simulation, DFA performed well when there was contrast in the underlying stock abundance (IncF and DecF). When conflicting surveys were introduced, though performance declined, DFA was still capable of accurately approximating the latent population trend from conflicting survey indices (Figures 3 and 4 and Supplementary Figures S1 and S2). Likewise, in the sandbar shark simulation, DFA performance remained relatively similar when up to three out of seven surveys experienced time-varying survey catchability (Figure 6 and Supplementary Figures S3 and S4).

There were, however, scenarios in which DFA performance broke down when conflicting indices were introduced. In the Atlantic sharpnose shark simulation, when there was no underlying contrast in the population size (ConstF), DFA performance declined considerably after the addition of survey conflict. In these instances, including a shift in survey catchability created a trend in the survey index that was not present in the population, and the DFA model could not decipher whether the signal was real or an artefact of the sampling. When the population trend was more complex (UF), conflicting indices hindered the ability of the DFA model to approximate the trend when three surveys were simulated, particularly when the surveys that underwent the shift had smaller CVs. When an incomplete, fourth survey was added (that did not experience a shift in the catchability coefficient), DFA performance improved marginally (Figure 4 and Supplementary Figures S1 and S2).

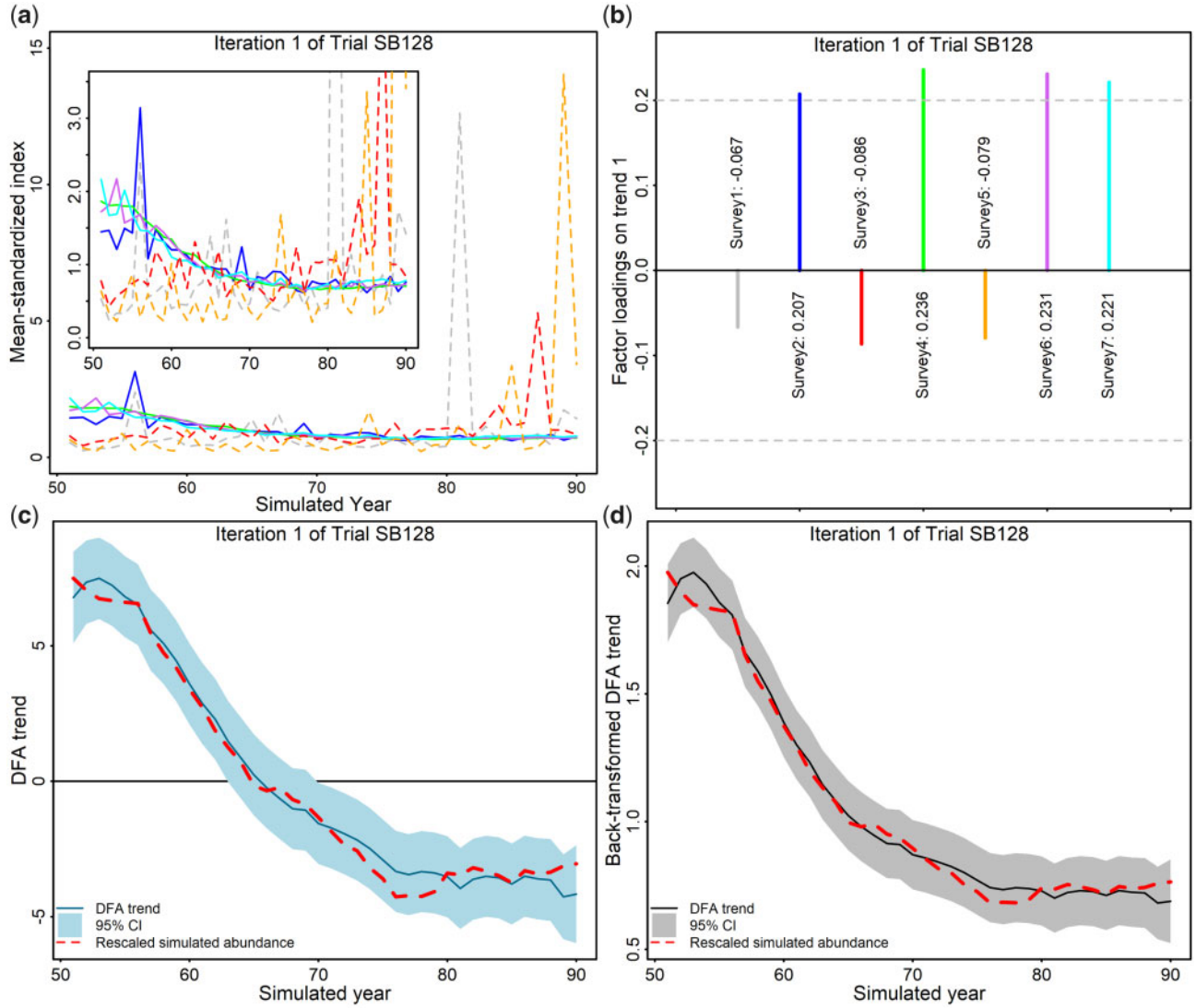

Figure 3. Example DFA model run for the first iteration of sandbar shark Trial SB128, including (a) input data as shown as mean-standardized survey indices with a zoomed inset, (b) corresponding factor loadings, denoting the strength of influence of the resulting DFA-predicted trend on each survey, (c) resulting DFA-predicted trend with 95% confidence intervals (CIs) in log space with 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 dashed lines).

Moreover, for both species, the direction of the shift in survey catchability relative to the direction of the population abundance change affected the ability of DFA to estimate an accurate trend. Consider the notable reduction in DFA performance within SN Trial UF_grad_2 (Figure 5). In this example, two of the three surveys underwent shifts in survey catchability. Survey 1 (S1) experienced a gradual increasing shift in q, while survey 3 (S3) underwent a gradual decreasing shift in q. The predominant trend of the underlying simulated population was decreasing, such that the shift in q for S1 was in the opposite direction of the population. Because the variability of S1 was smaller ($CV_1 = 0.3$) compared to S3 ($CV_3 = 0.7$), S1 was more heavily weighted in the DFA model. Thus, the DFA model placed a higher weight on an index that directly contradicted the latent population trend, resulting in poorer DFA model performance. Comparing this response of SN Trial UF_grad_2 to that of SN Trial UF_grad_3, the more heavily weighted survey experienced a shift in q that was in the same direction as the underlying population trend, and the resulting DFA predicted trend was more accurate (Figure 5).

In the sandbar shark simulation, where changes in survey catchability were explored in greater detail, DFA model performance was not greatly affected by increasing or decreasing patterns in catchability coefficient. However, the greatest reduction in DFA performance under the complete survey index scenario was observed when the survey with the smallest CV (S4, $CV_4 = 0.3$) was simulated under increasing q with two other surveys that experienced time-varying increases or decreases in q (Trials SB129 and SB138; Figure 7). Similar effects of equivalent magnitude were not observed when S4's q was altered in the same direction as the simulated population (e.g. compare Trials SB129 and

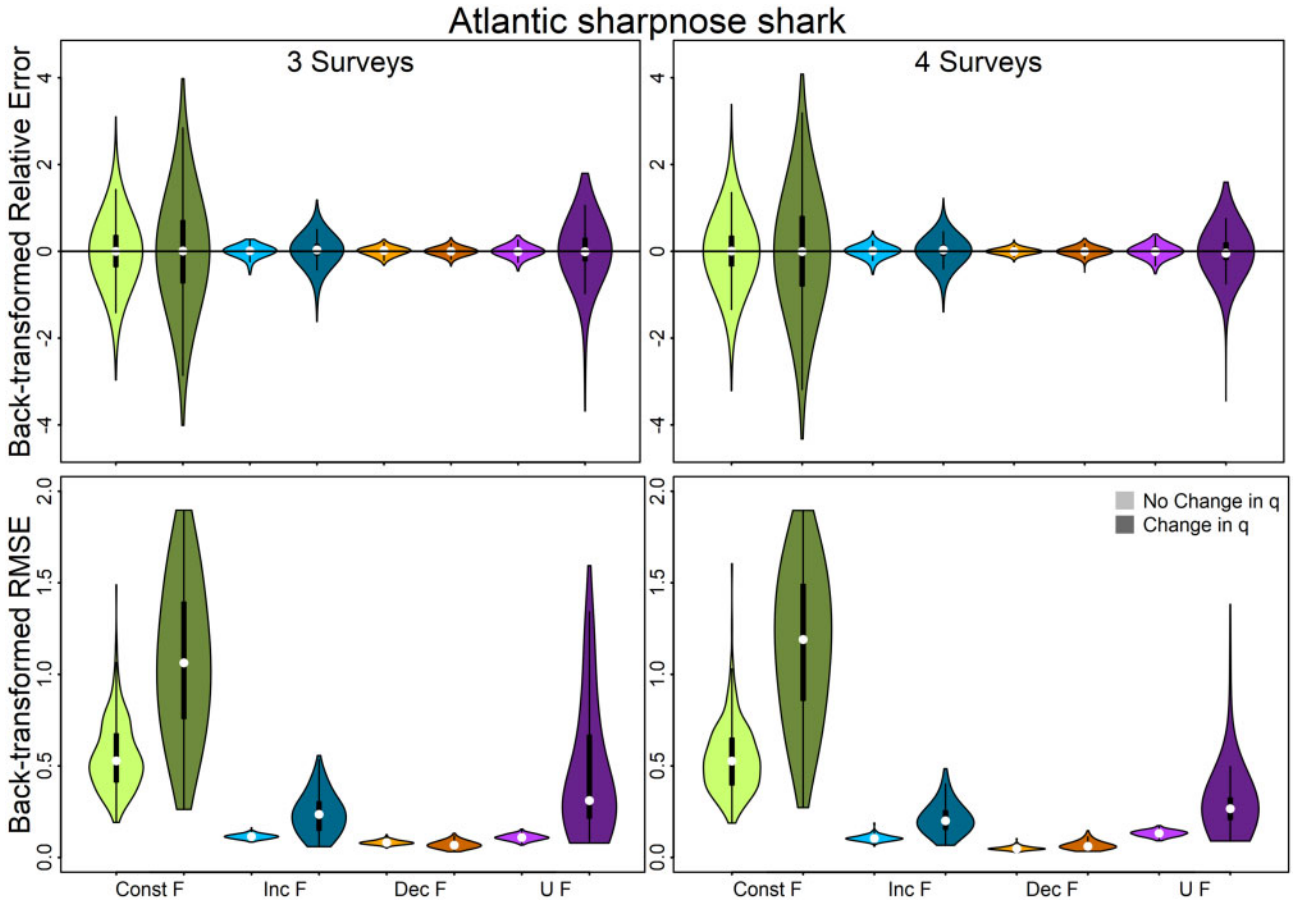

Figure 4. All simulation results for Atlantic sharpnose shark grouped by fishing mortality scenario, number of simulated surveys, and whether catchability was time-varying. Annual relative error and RMSE obtained by comparing standardized (*z*-scored) backtransformed DFA predicted trend to standardized, simulated abundance as applied to the Atlantic sharpnose shark. Spread in relative error should be narrow and RMSE should be low when DFA application was successful. All simulation results are grouped over survey CV and knife-edged or gradual shift in *q* (see Supplementary Figures S1 and S2 for results separated into one violin for each simulation scenario). See Table 1 for detailed description of each trial and for plotted scenario labels. Simulations in which three surveys were simulated are plotted in the left column, while simulations with four surveys are plotted in the right column. Violins are colour-coded into four categories (four violins per category), where each category represents a unique underlying population trend, as caused by variable fishing mortality (*F*) patterns: (1) constant *F*—constant population size, green; (2) increasing *F*—decreasing population size, blue; (3) decreasing *F*—increasing population size, orange; and (4) increasing then decreasing *F*—decreasing then increasing population size, purple. The first (lighter shaded) violins of each *F* scenario represent simulations for which no survey experienced a shift in catchability while the second (darker shaded) violins of each *F* scenario represent simulations for which two surveys experienced a shift in catchability. For full scenario-specific results, see the Supplementary figures.

SB138 to Trials SB132 and SB135, respectively; Figure 7). This response was also not distinguishable when surveys experienced missing years of data (see section "Missing data"; Figure 7).

We explored variable patterns of shifting survey catchability, either via knife-edged shifts, gradual shifts (Atlantic sharpnose simulation), or by maintaining consistent, time-varying increases or decreases of *q* throughout the simulated time-series (sandbar simulation). We found that in the Atlantic sharpnose shark simulation, the DFA model was generally able to more accurately predict the population trend when the shift in survey catchability was gradual compared to when it was knife-edged (Supplementary Figures S1 and S2).

Missing data

To more realistically approximate the quantity of data available in the sandbar shark simulation, we included the scenario in which several surveys were shorter than the full simulated time-series or had missing years of data, as emulated from the available information for the most recent sandbar shark assessment (Figure 1; SEDAR, 2017). The absence of complete survey data significantly reduced DFA performance (Figure 6). As expected, DFA was less capable of accurately estimating the trend of the underlying population when there was less (or no) information available to inform that year (early years in sandbar simulation). Consequently, the standard errors around the predicted DFA trend were much greater in earlier years when less information was available. The resulting DFA predicted trends generated under the same circumstances from missing data compared to complete data scenarios followed similar patterns, though the predictions from the missing data scenarios persistently underestimated early abundance, and were generally noisier than their smoother, complete data counterparts.

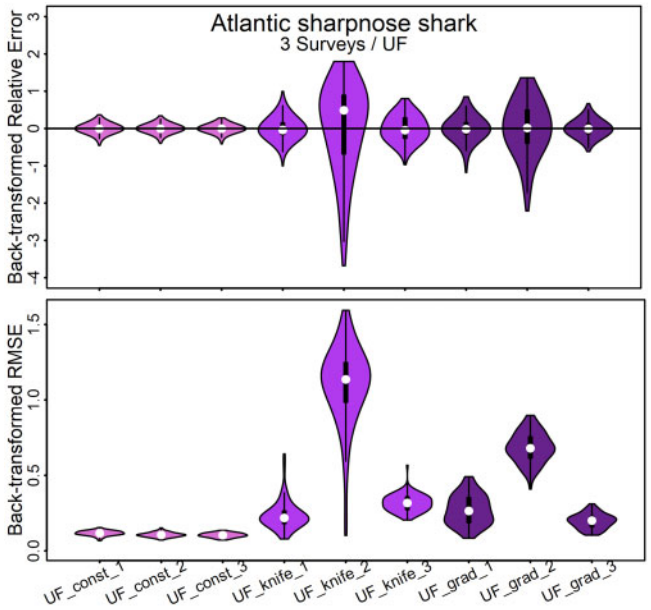

Figure 5. Scenario-specific Atlantic sharpnose shark simulation results for scenarios with three surveys simulated under the UF fishing mortality scenario. Annual relative error and RMSE obtained by comparing standardized (z-scored) backtransformed DFA predicted trend to standardized, simulated abundance as applied to the Atlantic sharpnose shark. Simulation results are presented for each simulation. Spread in relative error should be narrow and RMSE should be low when DFA application was successful. See Table 1 for more details on each scenario. For full scenario-specific results, see Supplementary figures.

The effect of conflicting indices on DFA performance in the case of missing data was greater than when data were complete. Predictably, when the most complete surveys (S1) underwent shifts in survey catchability, DFA performed more poorly than when the surveys with the fewest years of data (S7) underwent similar shifts (e.g. compare Trials SB28, 31, 34, and 37 to SB30, 33, 36, and 39 in Figure 7 and Supplementary Figure S4). The effect of the most complete survey (S1) undergoing shifts in q outweighed the effects of survey CV and directionality in the shift in q (Figure 7).

"All time-varying catchability" scenarios

As expected, when all surveys included a trend in catchability, the DFA was generally unable to recover the underlying stock abundance trend (Figures 6 and 7 and Supplementary Figures S3 and S4). In these scenarios, the DFA trend followed the shifting survey catchability signal rather than the desired signal of change in stock size. Nevertheless, the overall direction (increasing/decreasing) of the resulting DFA trend matched that of the stock abundance in the majority of simulated iterations. The DFA trends decreased over time, following the overall decreasing trend in stock size, in 84% of the "all time-varying catchability" scenario iterations.

The interaction between survey variability and the directionality of survey catchability trend was the main driver of the overall directionality of the DFA trend. When the survey with the smallest CV (S4) experienced increasing survey catchability, even when the majority of the surveys underwent decreasing survey catchability, the resulting DFA trend increased in 24.5% of iterations; this proportion was amplified in the missing data scenarios (44% DFA trends were increasing) compared to the complete data scenarios (5% DFA trends were increasing). On the other hand, when five surveys underwent an increase in survey catchability and two surveys (including S4) underwent a decrease in survey catchability, the resulting DFA trend followed the decreasing trend in the stock in 92.5% of iterations.

Discussion

Conflicting indices of abundance present a substantial challenge when trying to ascertain the true population trend over time. Not only do conflicting survey indices create confusion regarding trends in stock abundance, but also contradictory information is often passed to stock assessment models (Cortés *et al.*, 2015). Ultimately, the effect of conflicting survey indices in assessment models leads to uncertainty in the status of the stock (e.g. SEDAR, 2017) and has been deemed the "area of greatest concern" in a previous sandbar shark stock assessment (Hall, 2011; SEDAR, 2011).

When conflicts in indices of abundance are present, the disagreeing indices cannot all be simultaneously representative of total stock abundance (Schnute and Hilborn, 1993; Francis, 2017). Thus, we must interpret conflicting data in the context of sampling/observation and process error (Conn, 2010a, b; Maunder and Piner, 2017). Survey variability can be accounted for through sampling/observation error, while some variation in catchability and spatial distribution can be attributed to process error (Wilberg and Bence, 2006; Wilberg *et al.*, 2009; Conn, 2010a). DFA is a modified state-space approach, which explicitly considers observation and process uncertainty and is designed to accommodate short, nonstationary time-series (Zuur *et al.*, 2003b). Given the goal of accounting for and propagating uncertainty in fisheries science (e.g. Maunder, 2001; Maunder and Punt, 2013), DFA represents a practical approach to quantify known observation error (as calculated from CPUE standardization approaches; Maunder and Punt, 2004) and estimate additional

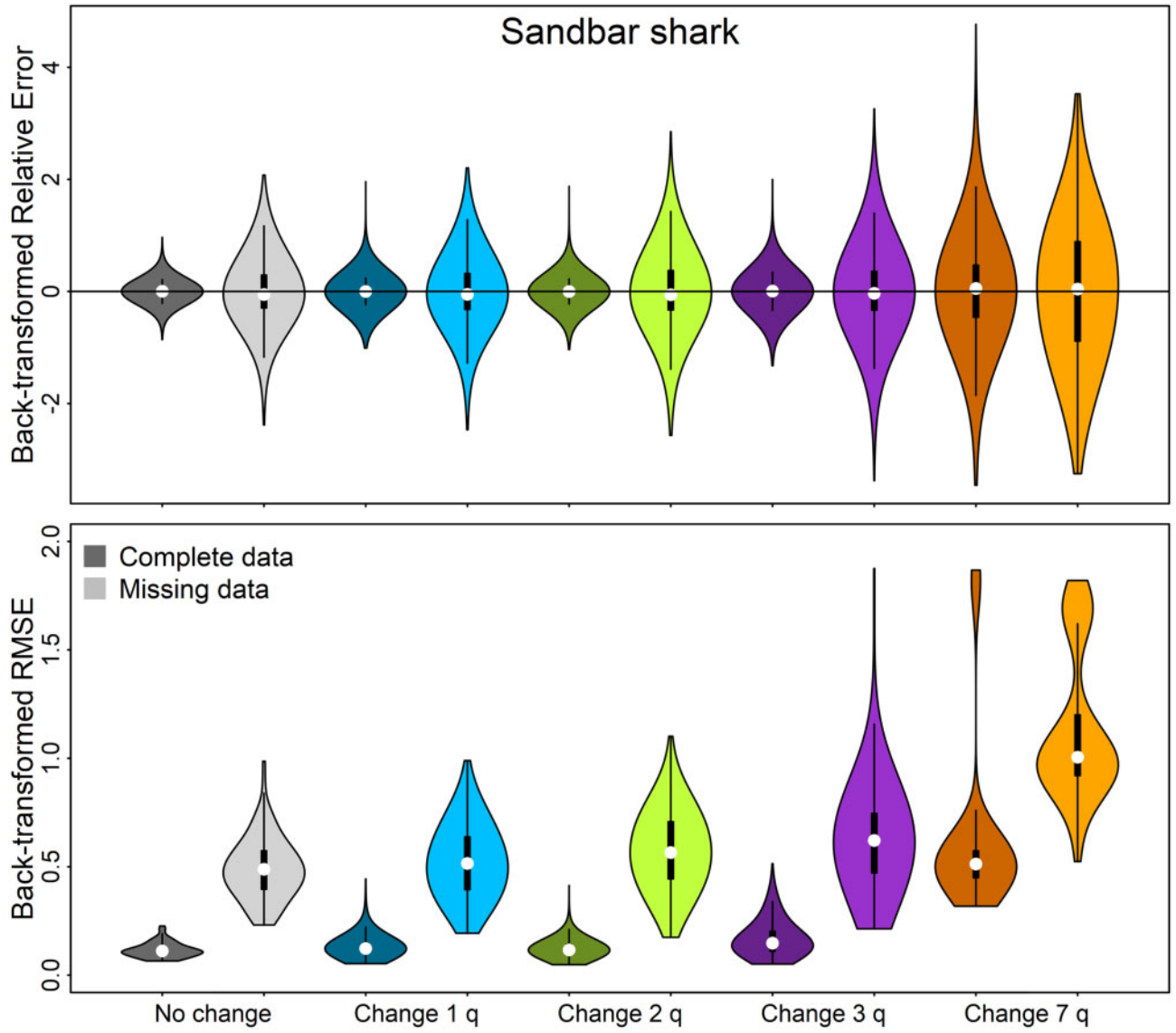

Figure 6. All simulation results for sandbar shark presented by the number of surveys with time-varying catchability and whether survey data were complete or contained missing years. Annual relative error and RMSE was obtained by comparing standardized (z-scored) backtransformed DFA predicted trend to standardized, simulated abundance as applied to the sandbar shark. Spread in relative error should be narrow and RMSE should be low when DFA application was successful. Results are presented for all trials grouped over directionality of shifting catchability. See Table 2 for detailed description of each trial and for respective plotting labels. Violins are paired; note a darker shaded violin is followed by a lighter shaded violin. The darker violin indicates simulation scenarios for which each survey had complete survey data, and lighter violins indicate simulation scenarios where all surveys experienced years of missing data. Violins are colour-coded based on the numbers of surveys that underwent a shift in catchability coefficient, *q*: (1) zero surveys, indicated by "No change", grey; (2) one survey, green; (3) two surveys, blue; (4) three surveys, purple; and (5) seven surveys, orange. For full scenario-specific simulation results, see Supplementary figures.

process error among survey indices outside the context of a stock assessment.

In our study, alterations in survey CVs resulted in slightly different estimated parameter variance, lending support to the proposition that DFA is properly tracking and propagating survey uncertainty. We chose CVs as the proxy for variance to reduce the impacts of index magnitude on input variance, though CV may not be appropriate for fishery-dependent indices that often have smaller CVs than surveys. When surveys with smaller CVs contradicted the underlying pattern in stock abundance, DFA performance decreased by more heavily weighting survey indices that experienced time-varying survey catchability. This result of unequal weighting could be potentially alleviated by allowing the DFA to internally estimate the parameters of the covariance matrix with a structure that assumes each survey variance is equal (thereby weighted equally). Alternatively, index-specific variances can also be estimated within the DFA model (Holmes *et al.*, 2014). The form of the factor loadings matrix can also be

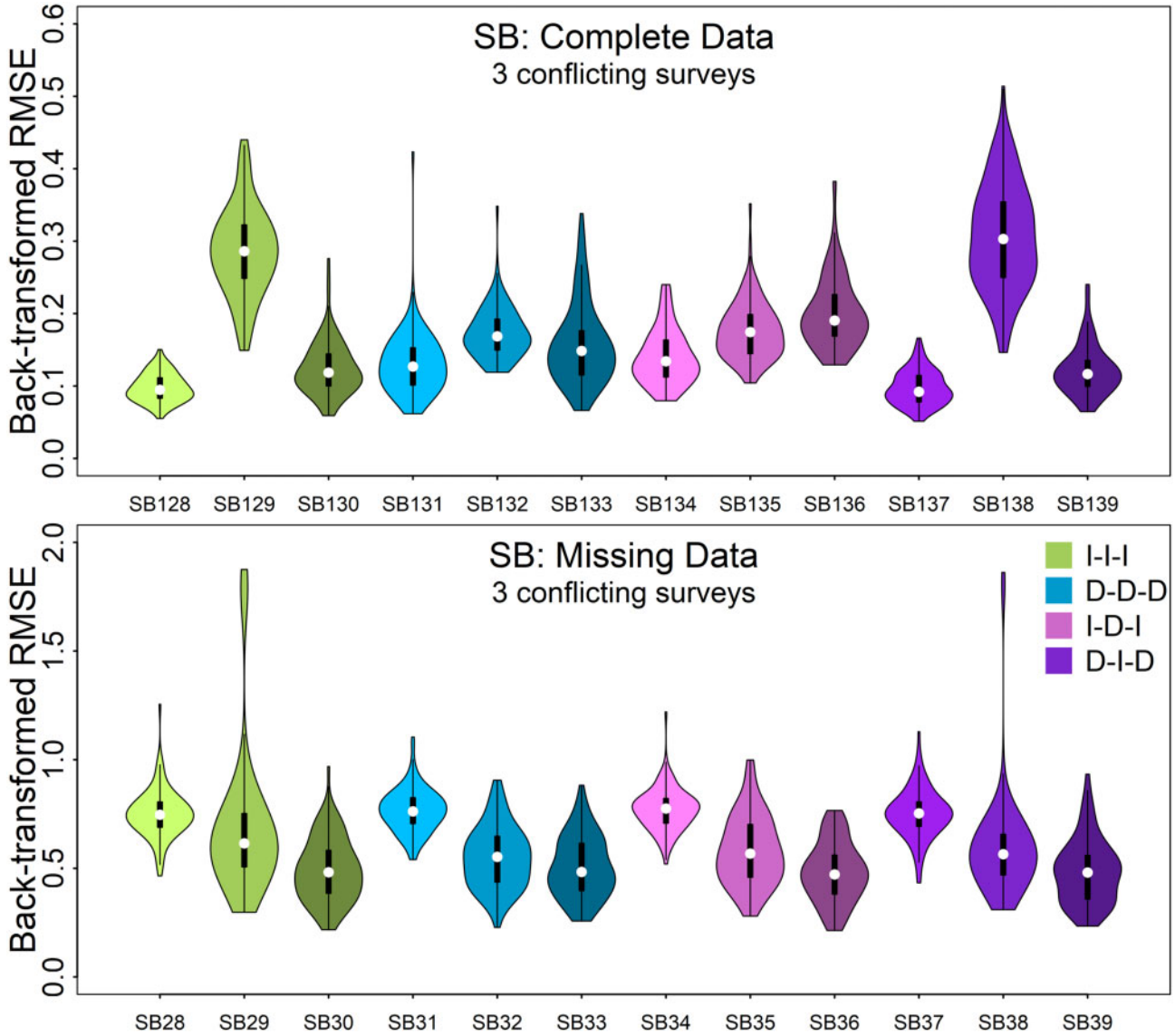

Figure 7. Scenario-specific sandbar shark simulation results for scenarios where three surveys underwent time-varying catchability. RMSE was obtained by comparing standardized (*z*-scored) backtransformed DFA predicted trend to standardized, simulated abundance as applied to the sandbar shark. Simulation results are presented for each simulation. Spread in RMSE should be low when DFA application was successful. Note the variable *y*-axes. Simulation scenarios are colour-coded based on the direction of the shift in catchability for each survey, where I indicates that the survey underwent an increase in the catchability coefficient and D signifies a decrease in the catchability coefficient (i.e. I-D-I, or pink-shaded violins, had the first and last time-varying surveys experience an increase in catchability coefficient, while the middle time-varying survey underwent a decrease in catchability coefficient). See Table 2 for more details on each scenario. For full scenario-specific results, see Supplementary figures.

modified to explicitly impose alternative weighting of input time-series (Holmes *et al.*, 2014). Other forms of index weighting by alternate measures (e.g. by spatial area covered, gear type, magnitude of catch) could be considered, and we suggest this as an area of further research.

We sought to explore the feasibility of using DFA as a tool to reconcile conflicting indices of relative abundance, while accounting for uncertainty, and thereby addressing concerns related to conflicts in survey abundance index trends. We chose to expand our simulation scenarios with the sandbar shark given the exploratory conceptual finding that after a series of survey indices were generated, life history strategy no longer influenced DFA's ability to reconcile time-series. Rather than repeating the same experiment with a different set of relative abundance indices, we expanded our simulation scenarios to more wholly characterize DFA performance in reconciling conflicting survey indices. We found that, in general, DFA performed fairly well as a method to reconcile conflicting indices across two species with unique life histories, data availability, variable survey selectivity, survey variability, and in the face of conflicting survey indices as simulated by changing survey catchability.

Given the results of the current simulation study, we compile our findings into the following list of recommendations when

using DFA as a method of reconciling conflicting indices of relative abundance:

1. Avoid using DFA when there is a lack of contrast in the underlying population abundance trend

As observed in the Atlantic sharpnose shark simulation, particularly in the presence of conflicting indices, DFA cannot accurately predict the underlying abundance trend when there is a lack of contrast in the stock size. Further, if there is a lack of contrast in all indices, then aggregating indices that are in agreement is less important. However, DFA performed well when the stock abundance was not constant over the analysed time-frame, even when indices were in conflict. If a DFA model produces a flat common trend and factor loadings close to or equal to 0, that does not necessarily mean that the DFA model did not work, but rather that the DFA model did not find a trend in the input time-series.

2. Use as many meaningful survey indices as possible

Our simulation results indicate that, despite the presence of conflicting survey indices, DFA performs better with additional survey inputs. While "no amount of statistical wizardry will remedy problems with poorly collected data" (Conn, 2010a, p. 118), it is challenging to determine which survey index or indices may be inappropriate and, therefore, should be excluded from analyses when limited data are available (Cortés, 2011). Likewise, identifying surveys that are representative of stock abundance and do not experience changing survey catchability *a priori* without expert knowledge of the system may also be impractical. DFA represents a tool to simplify discordant time-series and assess the relative importance of each input. Furthermore, DFA-estimated factor loadings indicate which survey indices most closely agree, which indices predominantly describe the resulting trend and, therefore, which indices may be considered less representative of stock-wide abundance.

Ultimately, because DFA is an averaging approach, it is clear that when all available indices are non-representative of the underlying stock abundance, the resulting DFA trend will not be accurate. Consider the "all time-varying catchability" scenarios in which all surveys conflicted to an unequal degree (i.e. five surveys experienced increasing/decreasing survey catchability, while the remaining two experienced decreasing/increasing survey catchability; Figures 6 and 7). Intuitively, if no survey index is a meaningful representation of stock abundance, then DFA is not to yield a trend that is reflective of stock abundance.

3. Use complete time-series, where possible

In the realistic scenario in which several indices are incomplete and contain missing values, DFA performance declines markedly. Accordingly, poorly informed years (early years in the sandbar shark simulation) are accompanied by substantial increases in uncertainty and relative error. Thus, under circumstances where missing data cannot be avoided, we encourage cautious interpretation of results. When input survey indices contain many missing values, interpreting small-scale noise is not advisable, particularly in years where fewer data are present. However, examining the broader tendency in the DFA trend may still prove useful for interpretation purposes (e.g. overall increasing or decreasing trend). Although rare, there were cases (~9/3600, excluding "all time-varying catchability" scenarios) in the missing data sandbar shark simulations in which the overall DFA predicted trend was increasing, in direct contrast with the decreasing trend of the simulated population. Analyses such as magnitudes or rates of decrease/increase may be inaccurate under DFA analyses in which there were many missing data points.

4. Carefully consider index transformations and error structure

Most survey indices are assumed to have a lognormal, multiplicative error structure, which is not consistent with the assumption of multivariate normal error assumed in the DFA model. While the assumption of normality is not of fundamental importance (Zuur *et al.*, 2003b; Zuur and Pierce, 2004), we demonstrate a technique of rescaling that includes the proper treatment of error. Multiplying each index by a vector of appropriately defined constants ensures that the properties of the time-series inputted into the DFA are consistent with those of a *z*-scored time-series (i.e. mean of 0 and standard deviation of 1). By log-transforming the indices of abundance, we transformed the lognormal error into a normal error structure, and by standardizing using the global standard deviation of the demeaned, descaled survey indices, we allowed for a way of back-transforming the resulting DFA trend out of log- and *z*-space. This process creates a single DFA predicted trend in abundance in arithmetic space with lognormal error. Note that the results from our rescaled approach are consistent with trends produced via a standard DFA model (log-transformed and *z*-scored indices prior to model run). Our rescaling protocol is not strictly necessary for the interpretation of conflicting survey indices but becomes more important when considering future analyses using the DFA-predicted trend.

Multiplying survey indices of relative abundance by a constant is comparable to redefining effort, such that the scale of the index increases or decreases. Failure to multiply each survey index by an appropriate vector of constants (**c**) results in inappropriately fitted and likely incorrect DFA results. 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). In our simulation, we were unable to identify a proper **c** vector for each iteration. Therefore, DFA performance would likely improve if a more appropriate **c** was adopted for each individual iteration.

5. Compare with other knowledge of stock trends

If the results of a DFA model suggest trends that are inconsistent with other pieces of available information and/or do not match prior understanding of the status of the population, then the results should be questioned.

Non-constant catchability

Although several factors that affect the catchability coefficient can be accounted for via CPUE standardization approaches (e.g. boat effect, fishing methodology, station/area effects; Maunder and Punt, 2004; Peterson *et al.*, 2017b), there are likely drivers of changes in catchability that cannot be explained in practice (Maunder *et al.*, 2006; Wilberg *et al.*, 2009). For example, it is likely that catchability is or will be changing in the future as a

result of climate-induced shifts in species distribution (e.g. a range expansion may result in increased availability of a target species to a survey, consequently increasing catchability or vice versa) and migratory timing (Karp *et al.*, 2019; Townhill *et al.*, 2019; O'Leary *et al.*, 2020). Physiological and behavioural responses to a dynamic biotic and abiotic environment will likely alter availability and encounter rate (Cheung *et al.*, 2012; Wittmann and Pörtner, 2013; Kotwicki *et al.*, 2015), potentially in unexpected ways. For example, under ocean acidification conditions, shark olfactory capacity will be impaired (Dixson *et al.*, 2015), which may reduce the attracting properties of a baited gear. Altered fisher behaviour in response to ecosystem and management changes compounded by fish behavioural changes (e.g. density-dependent and effort-dependent catchability; Hilborn and Walters, 1992; Wilberg *et al.*, 2009) have been shown to violate the assumption of constant catchability. Learning and behaviour alteration (Mitchell *et al.*, 2020) may further influence fishing dynamics over time (Guttridge *et al.*, 2009). We emphasize that whenever possible, factors that affect catchability should be accounted for within the CPUE standardization (Hilborn and Walters, 1992).

The ever-evolving nature of a nonstationary ecosystem may inherently influence catchability dynamics. As such, it is likely that catchability will change over time in most surveys, though the magnitude and/or directionality and pattern of time-varying catchability is particularly challenging to predict. For instance, it may be reasonable to assume that catchability varies around a constant coefficient, in which case we may assume that catchability is constant and allow the error term to capture the annual deviations from the mean catchability coefficient. In our simulation, we included more than fourfold changes in catchability in distinctive patterns to generate indices that conflicted throughout the simulation period. These shifts are consistent with the range of those estimated for other stocks (e.g. Wilberg *et al.*, 2009 and references therein). Whether all surveys are expected to undergo catchability shifts of equivalent magnitude in practice remains unclear and surely depends on the real-world system to which DFA would be applied. For example, in a system with two survey locations, catchability could increase in one location and decrease in the other if the population centre moved from the second location to the first. Situations like these may be expected for surveys with small spatial footprints under climate-change scenarios.

Nevertheless, given the multitude of, and complex interactions between, biological drivers of changing catchability, the goals of this study were not to hypothesize potential realistic scenarios or make inferences on which scenarios are more probable. Instead, we generated many abstract scenarios with realistic amounts of change in catchability to understand how DFA performs more conceptually. However, caution should be used in applying DFA to reconcile indices of abundance in situations with changes in catchability greater than we simulated. Further development of methods to detect large changes in catchability remains a high priority because many changes may be undetected with only conventional survey data (Wilberg *et al.*, 2009).

DFA approach

DFA is a consensus-type approach to data reconciliation. Alternative methods using spatio-temporal approaches (e.g. Thorson *et al.*, 2015a; Grüss and Thorson, 2019; Perretti and Thorson, 2019; O'Leary *et al.*, 2020; Thorson *et al.*, 2020) may be more appropriate when sufficient spatial data are available. If catchability changes are largely due to changes caused by availability of the stock to the survey, then spatial approaches have the benefit of mechanistically describing the cause of changing catchability. While DFA may not be the optimal statistical approach for identifying the underlying causes of conflicting indices of abundance in more data-rich stocks, environmental, climatic, and anthropogenic covariates can be included in a DFA model to infer potential causal factors (e.g. Buchheister *et al.*, 2016; Peterson *et al.*, 2017a). In addition, DFA can be used without specifying the mechanism for catchability change, which is both beneficial and a potential limitation.

DFA is a flexible approach with extensive options beyond those that we explored in the current study. For example, within the DFA modelling approach, users have the flexibility to account for covariation between survey indices by altering the structure of the observation error covariance matrix (e.g. Bers *et al.*, 2013; Colton *et al.*, 2014; Stachura *et al.*, 2014; Jorgensen *et al.*, 2016), incorporate broad-scale drivers of abundance in the form of covariates (e.g. Katara *et al.*, 2011; Bers *et al.*, 2013; Stachura *et al.*, 2014), and multiple common trends can be estimated (e.g. Bers *et al.*, 2013; Colton *et al.*, 2014; Jorgensen *et al.*, 2016), among other possibilities. Weights of input survey indices can be manually designated by fixing elements of the factor loadings matrix (Holmes *et al.*, 2014). Tools for the application of DFA within a Bayesian context (nwfsc-timeseries.github.io), as well as a spatial DFA variant (Thorson *et al.*, 2015b), have also been developed. Although we did not include all of these options (e.g. alternative forms of index weighting, survey covariability, spatial structure, or broad-scale climatic, environmental, or anthropogenic drivers) in the current study due to a simplified simulation framework, we highlight their presence to demonstrate the flexibility of DFA that may be required for various real-world implementation scenarios.

Study extensions

Azevedo *et al.* (2008) proposed utilization of DFA trends as an index of abundance within stock assessments. In unpublished companion research, we used the current simulation framework to explicitly test stock assessment performance with multiple conflicting survey indices vs. performance with a DFA trend inputted as relative abundance information. In this extension study, length composition data in the DFA assessment were weighted by DFA factor loadings and selectivity was estimated using more flexible, time-varying patterns (e.g. random-walk age-based selectivity with time-blocks). Though there is a general consensus that data be manipulated as little as possible (Maunder, 2001; Maunder and Punt, 2013; Methot and Wetzel, 2013), we consider the logical consistency of ensuring that survey indices are fulfilling their role within an assessment framework by acting as a representative measure of relative abundance.

Conclusion

We have shown that DFA can serve as a valuable tool for understanding and assessing the patterns of abundance of several fishes with many indices of relative abundance. However, DFA performance was relatively poor when no survey index is representative of stock abundance. Under such conditions, DFA was unable to provide accurate trends in abundance, as noted in our "all time-varying catchability" scenarios. Though this study focused on coastal sharks in the United States as catalyzed by the sandbar

shark stock assessment recommendations (Conn, 2010b; Cook, 2010; Hall, 2011; SEDAR, 2011, 2017), we found that the results were robust to the differences in life history strategy and data availability between a small coastal and a large coastal shark species. Consequently, this approach can be used for any fish stock that can be adequately surveyed with multiple indices of relative abundance, even across multiple selectivity patterns and when the assumption that indices are proportional to total abundance is violated in some cases. Fishes constantly cross geopolitical boundaries, resulting in multiple survey indices, and given relatively large observation and process errors, trends in those survey indices typically conflict. This study serves to provide guidance on use of DFA as an appropriate method to reconcile and interpret trends in fish abundance.

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

Data availability

There are no new data associated with this article.

Acknowledgements

We thank Andrew Scheld, Kyle Shertzer, Jim Thorson, and an anonymous reviewer for reviews of this manuscript. CDP was funded by the NMFS Sea Grant Population and Ecosystem Dynamics Fellowship (NA17OAR4170242). This paper is Contribution No. 3996 of the Virginia Institute of Marine Science, William & Mary.

References

Azevedo, M., Duarte, R., Cardador, 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: 1362–1369.

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.

Bers, A. V., Momo, F., Schloss, I. R., and Abele, D. 2013. Analysis of trends and sudden changes in long-term environmental data from King George Island (Antarctica): relationships between global climatic oscillations and local system response. Climatic Change, 116: 789–803.

Buchheister, A., Miller, T. J., Houde, E. D., Secor, D. H., and Latour, R. J. 2016. Spatial and temporal dynamics of Atlantic menhaden (*Brevoortia tyrannus*) recruitment in the Northwest Atlantic Ocean. ICES Journal of Marine Science, 73: 1147–1159.

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

Chen, C. S., Pierce, G. J., Wang, J., Robin, J. P., Poulard, J. C., Pereira, J., Zuur, A. F., *et al.* 2006. The apparent disappearance of *Loligo forbesi* from the south of its range in the 1990s: trends in *Loligo* spp. abundance in the northeast Atlantic and possible environmental influences. Fisheries Research, 78: 44–54.

Cheung, W. W. L., Pinnegar, J., Merino, G., Jones, M. C., and Barange, M. 2012. Review of climate change impacts on marine fisheries in the UK and Ireland. Aquatic Conservation: Marine and Freshwater Ecosystems, 22: 368–388.

Colton, A. R., Wilberg, M. J., Coles, V. J., and Miller, T. J. 2014. An evaluation of the synchronization in the dynamics of blue crab (*Callinectes sapidus*) populations in the western Atlantic. Fisheries Oceanography, 23: 132–146.

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.

Conrath, C. L., and Musick, J. A. 2012. Reproductive biology of elasmobranchs. *In* Biology of Sharks and Their Relatives, 2nd edn, pp. 291–311. Ed. by J. C. Carrier, J. A. Musick, and M. R. Heithaus. CRC Press, Boca Raton, FL. 672 pp.

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

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. Indian Ocean Tuna Commission, Working Party on Ecosystems and Bycatch, Document IOTC-2011-WPEB07-25. 11 pp.

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

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.

Dixson, D. L., Jennings, A. R., Atema, J., and Munday, P. L. 2015. Odor tracking in sharks is reduced under future ocean acidification conditions. Global Change Biology, 21: 1454–1462.

Ellis, J. R., Clarke, M. W., Cortés, E., Heessen, H. J. L., Apostolaki, P., Carlson, J. K., and Kulka, D. W. 2008. Management of elasmobranch fisheries in the North Atlantic. *In* Advances in Fisheries Science: 50 Years on from Beverton and Holt, pp. 184–228. Ed. by A. Payne, J. Cotter, and T. Potter. Blackwell Publishing, Oxford, UK. 568 pp.

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, W. B., III, 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.

Grubbs, R. D. 2010. Ontogenetic shifts in movements and habitat use. *In* Sharks and Their Relatives II. Biodiversity, Adaptive Physiology, and Conservation, pp. 319–350. Ed. by J. C. Carrier, J. A. Musick, and M. R. Heithaus. CRC Press, Boca Raton, FL. 672 pp.

Grüss, A., and Thorson, J. T. 2019. Developing spatio-temporal models using multiple data types for evaluating population trends and habitat usage. ICES Journal of Marine Science, 76: 1748–1761.

Guttridge, T. L., Myrberg, A. A., Porcher, I. F., Sims, D. W., and Krause, J. 2009. The role of learning in shark behaviour. Fish and Fisheries, 10: 450–469.

Hall, N. G. 2011. Desk review of stock assessments for sandbar, dusky, and blacknose shark prior to their consideration by the SEDAR 21 panel Review Workshop. 41 pp. http://sedarweb.org/cie-desk-reviewer-report-hall.

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. 585 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.

Holmes, E., Ward, E., and Wills, K. 2018. MARSS: Multivariate Autoregressive State-Space Modeling. R package version 3.10.10.

Jorgensen, J. C., Ward, E. J., Scheuerell, M. D., and Zabel, R. W. 2016. Assessing spatial covariance among time series of abundance. Ecology and Evolution, 6: 2472–2485.

Karp, M. A., Peterson, J. O., Lynch, P. D., Griffis, R. B., Adams, C. F., Arnold, W. S., Barnett, L. A. K., *et al.* 2019. Accounting for shifting distributions and changing productivity in the development of scientific advice for fishery management. ICES Journal of Marine Science, 76: 1305–1315.

Katara, I., Pierce, G. J., Illian, J., and Scott, B. E. 2011. Environmental drivers of the anchovy/sardine complex in the Eastern Mediterranean. Hydrobiologia, 670: 49–65.

Kotwicki, S., Horne, J. K., Punt, A. E., and Ianelli, J. N. 2015. Factors affecting the availability of walleye pollock to acoustic and bottom trawl survey gear. ICES Journal of Marine Science, 72: 1425–1439.

Latour, R. J., Gartland, J., and Bonzek, C. F. 2017. Spatiotemporal trends and drivers of fish condition in Chesapeake Bay. Marine Ecology Progress Series, 579: 1–17.

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

Mitchell, J. D., Schifiliti, M., Birt, M. J., Bond, T., McLean, D. L., Barnes, P. B., and Langlois, T. J. 2020. A novel experimental approach to investigate the potential for behavioural change in sharks in the context of depredation. Journal of Experimental Marine Biology and Ecology, 530–531: 151440.

O'Leary, C. A., Thorson, J. T., Ianelli, J. N., and Kotwicki, S. 2020. Adapting to climate-driven distribution shifts using model-based indices and age composition from multiple surveys in the walleye pollock (*Gadus chalcogrammus*) stock assessment. Fisheries Oceanography, 29: 541–557.

Ouellet, P., Savenkoff, C., Benoît, H. P., and Galbraith, P. S. 2016. A comparison of recent trends in demersal fish biomass and their potential drivers for three ecoregions of the Gulf of St Lawrence, Canada. ICES Journal of Marine Science, 73: 329–344.

Perretti, C. T., and Thorson, J. T. 2019. Spatio-temporal dynamics of summer flounder (*Paralichthys dentatus*) on the Northeast US shelf. Fisheries Research, 215: 62–68.

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

Peterson, C. D., Gartland, J., and Latour, R. J. 2017b. Novel use of hook timers to quantify changing catchability over soak time in longline surveys. Fisheries Research, 194: 99–111.

Pilling, G. M., Apostolaki, P., Failler, P., Floros, C., Large, P. A., Morales-Nin, B., Reglero, P., *et al.* 2008. Assessment and management of data-poor fisheries. *In* Advances in Fisheries Science: 50 Years on from Beverton and Holt, pp. 280–305. Ed. by A. Payne, J. Cotter, and T. Potter. Blackwell Publishing, Oxford, UK. 568 pp.

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.

Simpfendorfer, C. A., and Heupel, M. R. 2012. Assessing habitat use and movement. *In* Biology of Sharks and Their Relatives, 2nd edn, pp. 579–601. Ed. by J. C. Carrier, J. A. Musick, and M. R. Heithaus. CRC Press, Boca Raton, FL. 672 pp.

SouthEast Data, Assessment, and Review (SEDAR). 2011. SEDAR 21 HMS Sandbar Shark Stock Assessment Report. http://sedarweb.org/docs/sar/Sandbar_SAR.pdf.

SouthEast Data, Assessment, and Review (SEDAR). 2013. SEDAR 34 HMS Atlantic Sharpnose Shark Stock Assessment Report. http://sedarweb.org/docs/sar/S34_ATSH_SAR.pdf.

SouthEast Data, Assessment, and Review (SEDAR). 2017. SEDAR 54 HMS Sandbar Shark Stock Assessment Report. http://sedarweb.org/docs/sar/S54_Final_SAR_with_exec_summary.pdf.

Stachura, M. M., Mantua, N. J., and Scheuerell, M. D. 2014. Oceanographic influences on patterns in North Pacific salmon abundance. Canadian Journal of Fisheries and Aquatic Sciences, 71: 226–235.

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.

Tam, Y.-K., Ni, I. H., Yau, C., Yan, M.-Y., Chan, W.-S., Chan, S.-M., and Lu, H.-J. 2013. Tracking the changes of a fish community following a megascale reclamation and ensuing mitigation measures. ICES Journal of Marine Science, 70: 1206–1219.

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.

Thorson, J. T., Ciannelli, L., and Litzow, M. A. 2020. Defining indices of ecosystem variability using biological samples of fish communities: a generalization of empirical orthogonal functions. Progress in Oceanography, 181: 102244.

Thorson, J. T., Scheuerell, M. D., Shelton, A. O., See, K. E., Skaug, H. J., and Kristensen, K. 2015b. Spatial factor analysis: a new tool for estimating joint species distributions and correlations in species range. Methods in Ecology and Evolution, 6: 627–637.

Thorson, J. T., Shelton, A. O., Ward, E. J., and Skaug, H. J. 2015a. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES Journal of Marine Science, 72: 1297–1310.

Townhill, B. L., Radford, Z., Pecl, G., Putten, I., Pinnegar, J. K., and Hyder, K. 2019. Marine recreational fishing and the implications of climate change. Fish and Fisheries, 20: 977–992.

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. 2009. Incorporating time-varying catchability into population dynamic stock assessment models. Reviews in Fisheries Science, 18: 7–24.

Wittmann, A. C., and Pörtner, H.-O. 2013. Sensitivities of extant animal taxa to ocean acidification. Nature Climate Change, 3: 995–1001.

Zuur, A. F., Fryer, R. J., Jolliffe, I. T., Dekker, R., and Beukema, J. J. 2003a. Estimating common trends in multivariate time series using dynamic factor analysis. Environmetrics, 14: 665–685.

Zuur, A. F., and Pierce, G. J. 2004. Common trends in northeast Atlantic squid time series. Journal of Sea Research, 52: 57–72.

Zuur, A. F., Tuck, I. D., and Bailey, N. 2003b. Dynamic factor analysis to estimate common trends in fisheries time series. Canadian Journal of Fisheries and Aquatic Sciences, 60: 542–552.

Handling editor: Emory Anderson