ICES Journal of Marine Science

ICES Journal of Marine Science (2019), 76(6), 1748–1761. doi:10.1093/icesjms/fsz075

Original Article

Developing spatio-temporal models using multiple data types for evaluating population trends and habitat usage

Arnaud Grüss[1]* and James T. Thorson[2]

[1]School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98105-5020, USA
[2]Habitat and Ecosystem Process Research Program, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, 7600 Sand Point Way N.E., Seattle, WA 98115, USA

*Corresponding author: tel: +1 305 606 5696; e-mail: gruss.arnaud@gmail.com

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

Received 30 January 2019; revised 27 March 2019; accepted 29 March 2019; advance access publication 25 April 2019.

Spatio-temporal models have become key tools for evaluating population trends and habitat usage. We developed a spatio-temporal modelling framework employing a combination of encounter/non-encounter, count, and biomass data collected by different monitoring programs ("combined data"). The three data types are predicted using a computationally efficient approximation to a compound Poisson-gamma process. We fitted spatio-temporal models to combined data for Gulf of Mexico (GOM) red snapper (*Lutjanus campechanus*) for 2006–2014. These spatio-temporal models provided insights into GOM red snapper spatial distribution patterns, which we corroborated by comparing to past predictions generated using only encounter/non-encounter data. However, relying on biomass and count data in addition to encounter/non-encounter data also allowed us to reconstruct biomass trends for GOM red snapper and to examine patterns of distribution shifts and range expansion/contraction for this population for the first time. Moreover, combining multiple data types improved the precision of reconstructed population trends and some variables quantifying habitat usage. Finally, scenarios and simulation experiments conditioned upon red snapper data showed that the improvement in fitting to combined data is greater when biomass data for the study population are lacking for an entire subregion and, to a lesser extent, for an entire time period (e.g. in early years).

Keywords: data-integrated models, habitat assessments, habitat usage, population assessments, population trends, red snapper (*Lutjanus campechanus*), simulation experiment, spatio-temporal models, state-space models.

Introduction

Ecologists are tasked with conducting population and habitat assessments for supporting resource management, using the best available data and models. For example, terrestrial ecologists carry out assessments of red deer (*Cervus elaphus*) populations to inform culling strategies in Scotland (Trenkel *et al.*, 2000) and determine the spatial abundance patterns of the invasive barred owl (*Strix varia*) population to guide spatial conservation plans in the Pacific Northwest, United States (Rossman *et al.*, 2016). Another example is that of US marine ecologists who assist resource managers implementing the Magnuson-Stevens Fishery Conservation and Management Act (MSRA, 2006). The Magnuson-Stevens Fishery Conservation and Management Act requires the regular update of harvest limits based on the population assessments of the target species, commonly termed "stock assessments" (Federal Register, 2008); and the designation of essential fish habitat based on scientific information on the spatial distribution patterns of juveniles and spawners of the species of interest (Rosenberg *et al.*, 2000).

State-space models, which consider both hidden population dynamics and uncertainties in the observation process (Schnute, 1994), are key tools for supporting population and habitat assessments. In the terrestrial world, they have been used, for instance, to understand the population trends and spatial distribution patterns of barred owl in the Pacific Northwest (Rossman *et al.*, 2016; Zipkin *et al.*, 2017). In the marine world, spatio-temporal models, state-space models that account for spatial and spatio-temporal structure at a fine scale, are increasingly being employed

to standardize sample count and biomass data (Berg *et al.*, 2014; Thorson *et al.*, 2015; Cao *et al.*, 2017; Grüss, Walter, *et al.*, 2019) and reconstruct population age/length structures from subsampling data (Thorson and Haltuch, 2018) for informing population assessments; and to map nursery and spawning habitats (Kai *et al.*, 2017; Grüss, Biggs, *et al.*, 2018) and estimate patterns of distribution shifts and range expansion/contraction (Thorson, Pinsky, *et al.*, 2016; Thorson, Rindorf, *et al.*, 2016) for assisting habitat assessments.

One major challenge for terrestrial and marine scientists is that individual datasets often do not provide sufficient information to conduct satisfactory ecological investigations (Schaub and Abadi, 2011; Grüss *et al.*, 2017). To remedy this issue in the terrestrial world, integrated population models (IPMs) have been designed (Schaub and Abadi, 2011). IPMs are state-space models that carry out a joint analysis of multiple datasets on different quantities. Inference with IPMs relies on the joint likelihood, which can be constructed only if the individuals included in the different datasets can be assumed to have the same demography and if the likelihoods for the individual datasets have parameters in common (Maunder, 2004; Schaub and Abadi, 2011). The advantages of using IPMs instead of models fitted to individual datasets on particular quantities are manifold. Importantly, with IPMs, more of the information provided by individual datasets can be exploited, resulting in the estimation of a larger number of demographic quantities, as well as in more precise estimates for individual demographic quantities (Brooks *et al.*, 2004; Schaub *et al.*, 2007; Abadi *et al.*, 2012; Zipkin *et al.*, 2017). Simulation experiments also found that employing IPMs led to more accurate estimates of population abundance and demographic rates (Dorazio, 2014; Zipkin *et al.*, 2017). IPMs have long focused on the joint analysis of capture–recapture data and other data types such as telemetry data and fecundity estimates (Lebreton *et al.*, 1995; Abadi *et al.*, 2012; Wilson *et al.*, 2016). More recently, IPMs referred to as "dynamic N-mixture models" have considered a combination of multiple "unmarked" data types such as detections/non-detections (i.e. encounters/non-encounters) and quadrat counts (Dail and Madsen, 2011; Dorazio, 2014; Zipkin *et al.*, 2017).

In the marine world, the spatio-temporal models developed for supporting population and habitat assessments have, in general, analysed individual datasets separately (Berg *et al.*, 2014; Thorson, Rindorf, *et al.*, 2016; Kai *et al.*, 2017; Thorson *et al.*, 2017). Exceptions to this general pattern include spatio-temporal modelling studies that combined encounter/non-encounter (Grüss *et al.*, 2017, Grüss, Biggs, *et al.*, 2018; Grüss, Drexler, *et al.*, 2018; Grüss, Perryman, *et al.*, 2018; Grüss, Thorson, *et al.*, 2018), count (Runnebaum *et al.*, 2018) or biomass (Dolder *et al.*, 2018; Perretti and Thorson, 2019) data collected by different monitoring programs. In all cases, differences in detection probability using samples from different sampling programs were accounted for through a coefficient representing differences in capture efficiency. For example, Grüss *et al.* (2017) and Grüss, Perryman, *et al.* (2018) compiled a large monitoring database for the US Gulf of Mexico (GOM) storing the encounter/non-encounter data collected by different fisheries-independent and fisheries-dependent programs employing random sampling schemes, and the authors then fitted spatio-temporal models to the large monitoring database to be able to map the spatial distributions of fish and invertebrate species groups, species, and life stages. Without this large endeavour, it would not have been possible to map the spatial distributions of many species and life stages at the scale of the entire US GOM, because the spatial footprint of individual monitoring programs would have been insufficiently small (Grüss *et al.*, 2017; Grüss, Perryman, *et al.*, 2018). However, a more comprehensive use of the information provided by monitoring data would substantially benefit population and habitat assessments. More specifically, fitting spatio-temporal models to a combination of encounter/non-encounter, count and biomass data collected by different monitoring programs would allow for the reconstruction of population trends to inform population assessments, and for the estimation of patterns of distribution shifts and range expansion/contraction to assist habitat assessments (Grüss, Perryman, *et al.*, 2018).

In this study, we draw inspiration from the IPMs designed for terrestrial studies to develop a spatio-temporal modelling framework using a combination of encounter/non-encounter, count, and biomass data (henceforth "combined data"). Our primary goal is to explore the impacts of employing combined data (i.e. a "data-integrated model") vs. biomass-only data (i.e. a conventional model in fisheries analyses) for informing population and habitat assessments. In the following, we specify the state-process and observation components of our spatio-temporal models and then describe parameter estimation. Next, we examine the predictions, precision, accuracy, error, and confidence interval coverage of spatio-temporal models fitted to combined data for GOM red snapper (*Lutjanus campechanus*) for the period 2006–2014, in three steps. First, we conduct a comparison of the population trends and patterns of distribution shifts and range expansion/contraction and their uncertainty predicted by a spatio-temporal model fitted to combined vs. biomass-only data. Second, because many fish and invertebrate populations (hereafter simply referred to as "fish populations") are not adequately sampled compared to GOM red snapper, we repeat this comparison under two "extreme" scenarios: a "spatial" scenario, where all the biomass data except for 98% of those for the northwestern Gulf of Mexico (NWGOM) are considered; and a "temporal" scenario, where all the biomass data except for 98% of those for the period 2006–2008 are considered. Third, we use a simulation experiment to evaluate the accuracy, error, and confidence interval coverage of the population trends predicted by the spatio-temporal model fitted to combined data, when biomass data are either collected or not collected in the NWGOM or when biomass data are either collected or not collected in early years.

Methods

Model specifications

Our model is a spatio-temporal Poisson-link delta model (Thorson, 2018a), where state variables include both the numbers-density of the species of interest, $n(s,t)$, and the biomass-per-individual of this species, $w(s,t)$, at each site s and year t. The product of these two quantities gives the biomass-density of the species of interest, $d(s,t)$, at each site s and year t.

In the marine environment, it is common to sample biomasses, employing, e.g. trawls, seines, or traps (Grüss, Perryman, *et al.*, 2018). Many monitoring programs also collect count data using, e.g. longlines, vertical lines, or video cameras. Monitoring programs that specifically aim to record encounters/non-encounters are rare since any sampling protocol usually allows one to obtain at least counts (e.g. video cameras, visual censuses). However, the counts or biomasses collected by monitoring programs are often converted into encounters/non-encounters for fitting species

distribution models to a blending of data coming from different surveys (Grüss *et al.*, 2017; Pirtle *et al.*, 2017; Grüss, Perryman, *et al.*, 2018), or for developing conventional two-stage, delta models (combining, e.g. a binomial model fitted to encounter/non-encounter data and a quasi-Poisson model fitted to count data; Grüss *et al.*, 2014, 2016).

Encounter/non-encounter data R can take the value 0 (not encountered) or 1 (encountered). When fitting to encounter/non-encounter data, we assume that the spatial distribution of individuals in the neighbourhood of sampling is random, such that the probability to encounter at least one individual follows a Poisson distribution with intensity equal to local numbers-densities times area sampled. This results in a Bernoulli distribution with an encounter probability $p(i)$ using a complementary log–log link function given log numbers-density, $\log(n(s, t))$:

$$\begin{aligned} R &\sim \text{Bernoulli}\Big(p(i)\Big) \\ p(i) &= 1 - \exp\Big(-a_i n(s_i, t_i)\Big), \end{aligned} \tag{1}$$

where a_i is the area sampled for sample i, in km^2 (e.g. the area swept if sample i was collected by a trawl survey).

Count data C can take any positive integer. When fitting to count data, we again assume that individuals are randomly distributed in the proximity of sampling. This results in a Poisson distribution for count data, with intensity equal to local numbers-density times area sampled. However, count samples are frequently overdispersed relative to predictions of local density. We, therefore, expand our model to estimate the magnitude of overdispersion using a lognormal-Poisson distribution:

$$\begin{aligned} C &\sim \text{Poisson}\Big(\lambda(i)\Big) \\ \lambda(i) &= a_i n(s_i, t_i) \times e^{\delta(i)}, \end{aligned} \tag{2}$$

where $\delta(i) \sim \text{Normal}(0, \sigma^2_{\text{obs}})$ is a random effect representing normally distributed overdispersion with variance σ^2_{obs}.

Biomass-sampling data B can take any non-negative real number. When fitting to biomass-sampling data, we employ a Poisson-link delta model and follow Thorson (2018a) in defining a probability of encounter $p(i)$ [Equation (1)] and the expected biomass given that the species is encountered r_i (subsequently called "positive catch rate"). The product of these two variables is equal to expected biomass-density, $\mathbb{E}(d) = p \times r$, and is also equal to the product of numbers-density n and biomass-per-individual w. We can, therefore, calculate positive catch rate as:

$$r(i) = \frac{n(s_i, t_i)}{p(i)} \; w(s_i, t_i). \tag{3}$$

We can, then, define a delta model for biomass-samples:

$$\Pr(b(i) = B) = \begin{cases} 1 - p(i) & \text{if } B = 0 \\ p(i) \times g(B|r(i), \sigma^2_{\text{obs}}) & \text{if } B > 0, \end{cases} \tag{4}$$

where $g\big(B|r(i), \sigma^2_{\text{obs}}\big)$ is the gamma probability density function for unexplained variation in $r(i)$; and σ^2_{obs} is residual biomass sampling variation.

Thus, using a Poisson-link delta model allows the likelihoods for encounter/non-encounter, count and biomass datasets to have parameters in common [Equations (1)–(4)]. As a result, when multiple data types are provided to our spatio-temporal model, it is straightforward to calculate the likelihood of the observation-process equations as the product of the likelihoods for encounter/non-encounter, count, and biomass datasets (Schaub and Abadi, 2011; Dorazio, 2014).

Fitting biomass, count, and/or encounter/non-encounter data to our spatio-temporal model allows for the estimation of numbers-density and biomass-per-individual:

$$\begin{aligned} \log\Big(n(s_i, t_i)\Big) &= \beta_n(t_i) + \omega_n(s_i) + \varepsilon_n(s_i, t_i) + \sum_{m=1}^{n_m} \gamma_m G(i, m) \\ \log\Big(w(s_i, t_i)\Big) &= \beta_w(t_i) + \omega_w(s_i) + \varepsilon_w(s_i, t_i), \end{aligned} \tag{5}$$

where $\beta_n(t_i)$ and $\beta_w(t_i)$ are intercepts for the year t_i in which sample i was collected, which are both estimated as fixed effects; $\omega_n(s_i)$ and $\omega_w(s_i)$ represent spatial variation and are both estimated as random effects; $\varepsilon_n(s_i, t_i)$ and $\varepsilon_w(s_i, t_i)$ represent spatio-temporal variation and are both estimated as random effects; n_m is the total number of monitoring programs considered; and $\sum_{m=1}^{n_m} \gamma_m G(i, m)$ is the effect of monitoring programs on the expected number of individuals sampled, which is turned off if only one monitoring program provides data to the model (i.e. if $n_m = 1$). The design matrix $G(i, m)$ is such that $G(i, m)$ is 1 for the monitoring program m that collected sample i and 0 otherwise, and the monitoring program effect γ_m is such that $\gamma_m = 0$ for the monitoring program m associated with the largest sample size to allow for the identifiability of all β_n parameters. We treat the effect of monitoring programs as fixed here, although future studies could treat it as random via the implementation of restricted maximum likelihood (Grüss *et al.*, 2017; Grüss, Perryman, *et al.*, 2018). Moreover, Equation (5) could include environmental covariates, but we do not consider this option here and leave it for future research.

The spatial and spatio-temporal variation terms are all assumed to follow a multivariate normal distribution:

$$\begin{aligned} \boldsymbol{\omega}_{\text{n}} &\sim \text{MVN}\Big(0, \sigma^2_{n\omega}\mathbf{R}(\kappa_n)\Big) \\ \boldsymbol{\varepsilon}_{\text{n(t)}} &\sim \text{MVN}\Big(0, \sigma^2_{n\varepsilon}\mathbf{R}(\kappa_n)\Big) \\ \boldsymbol{\omega}_{\text{w}} &\sim \text{MVN}\Big(0, \sigma^2_{w\omega}\mathbf{R}(\kappa_w)\Big) \\ \boldsymbol{\varepsilon}_{\text{w(t)}} &\sim \text{MVN}(0, \sigma^2_{w\varepsilon}\mathbf{R}(\kappa_w)), \end{aligned} \tag{6}$$

where $\mathbf{R}(\kappa_n)$ is the correlation among sites as a function of decorrelation distance κ_n; $\mathbf{R}(\kappa_w)$ is the correlation among sites as a function of decorrelation distance κ_w; $\sigma^2_{n\omega}$ and $\sigma^2_{w\omega}$ are the estimated pointwise variances of the spatial variation in numbers-density and biomass-per-individual, respectively; and $\sigma^2_{n\varepsilon}$ and $\sigma^2_{w\varepsilon}$ are the estimated pointwise variances of the spatio-temporal variation in numbers-density and biomass-per-individual, respectively. All spatial and spatio-temporal variation terms are estimated using Gaussian Markov random fields (Thorson *et al.*, 2015).

Parameter estimation

As is the case for any state-space model, the joint likelihood of our spatio-temporal model fitted to individual data types (encounter/non-encounter, count, or biomass data) is the product of

the likelihoods of the state-process and observation-process equations (Maunder, 2004). When our model is fitted to multiple types of data simultaneously, the likelihoods for encounter/non-encounter, count, and biomass datasets have parameters in common [Equations (1)–(4)], so that the likelihood of the observation-process equations is the product of the likelihoods for encounter/non-encounter, count, and biomass datasets (Schaub and Abadi, 2011; Dorazio, 2014). However, this is allowable only if the encounter/non-encounter, count, and biomass datasets can be assumed to be statistically independent and if the individuals included in these different datasets can be assumed to belong to the same population (Maunder, 2004; Schaub and Abadi, 2011; Dorazio, 2014). These two conditions are usually fulfilled by the encounter/non-encounter, count, and biomass datasets collected in the same marine region using random sampling schemes, as the datasets for GOM red snapper employed later in this study.

To estimate the parameters of our spatio-temporal model, we used R package "VAST" (Thorson, 2019), whose code and associated materials are publicly available online (https://github.com/James-Thorson/VAST). The estimation of fixed effects in VAST is accomplished via the identification of the parameter values that maximize the joint likelihood. First, VAST employs the Laplace approximation implemented by R package "TMB" (Kristensen *et al.*, 2016) to compute the marginal likelihood by approximating the integral across all random effects. By employing automatic differentiation, TMB efficiently computes the matrix of second derivatives (which the Laplace approximation uses) and the gradient of the Laplace approximation (which is used when fixed effects are maximized). All random effects are predicted by TMB through the maximization of the joint likelihood given the maximum likelihood estimates (MLEs) of fixed effects. We employed Thorson and Kristensen (2016)'s bias-correction estimator to correct for the "retransformation bias" when predicting any derived quantity involving a nonlinear transformation of random effects. We also used the generalized delta method implemented in TMB to compute the standard errors (SEs) of all fixed and random effects and the SEs of derived quantities (Kass and Steffey, 1989).

Demonstration for GOM red snapper

We demonstrate our spatio-temporal model by applying it to red snapper, a socio-economically important species in the US GOM (Figure 1). GOM red snapper supports a multi-billion dollar recreational fishing industry, as well as a very large commercial fishery, which has historically been one of the GOM fisheries with the largest landings (NMFS, 2017). All the data we employed for GOM red snapper were collected between 2006 and 2014 by fisheries-independent surveys using random sampling schemes; the biomass data came from the SEAMAP Groundfish Trawl Survey ("TRAWL") dataset (Rester, 2017), the count data from the National Marine Fisheries Service (NMFS) Pelagic Acoustic Trawl Survey ("PELACTR") dataset (Pollack and Ingram, 2014), and the encounter/non-encounter data from the NMFS Red Snapper/Shark Bottom Longline Survey ("BLL") dataset (Henwood *et al.*, 2006). All surveys and datasets are detailed in Supplementary Appendix S1.

First, to gauge the population trends predicted by our spatio-temporal modelling framework, we examined the indices of relative biomass (henceforth "indices") and their SEs predicted by a spatio-temporal model fitted to the combined dataset vs. individual datasets. We also compared the indices predicted by the spatio-temporal models fitted to combined and biomass-only data to the biomass estimates predicted by the latest GOM red snapper population assessment (SEDAR 52, 2018). Then, to gauge the patterns of distribution shifts and range expansion/contraction predicted by our spatio-temporal modelling framework, we examined the eastward and northward centres of gravity (COGs) and effective areas occupied and the SEs of these quantities predicted by a spatio-temporal model fitted to the combined dataset vs. individual datasets (see below for calculation). We refer to this scenario where we use all of the original monitoring data for GOM red snapper as the "base scenario."

Next, because many fish populations are not adequately sampled compared to GOM red snapper, we repeated the comparison of the predictions of the spatio-temporal models fitted to combined vs. biomass-only data under two "extreme" scenarios: a "spatial" scenario, where all the biomass data except for 98% of those for the NWGOM were considered; and a "temporal" scenario, where all the biomass data except for 98% of those for the period 2006–2008 were considered. Given that the spatial patterns of biomass and exploitation of many populations differ greatly over space (e.g. some GOM snapper and shrimp populations; Grüss, Perryman, *et al.*, 2018), a severe lack or an absence of biomass-sampling data for an entire subregion can lead to the reconstruction of very inaccurate population trends if the analysts rely only on biomass-sampling data. Moreover, a severe lack or an absence of biomass-sampling data for the early years of a fishery can result in inappropriate baselines upon which to gauge the status of populations if the analysts make use of only biomass-sampling data (Pauly, 1995).

For all analyses and the simulation experiment described below, the value of all spatial and spatio-temporal variation terms defined over a fixed spatial domain Ω ($s \in \Omega$) was approximated as being piecewise constant, for computational reasons. We specified 200 "knots" to approximate all the spatial and spatio-temporal variation terms defined over domain Ω, such that the spatio-temporal models tracked the value of these terms at each knot (Figure 2; Shelton *et al.*, 2014). Thus, the value of a spatial or spatio-temporal variation term at a given site was established from the value of the term at the knot that was the closest to that site. Knot location was decided via the application of a *k*-means algorithm to the locations of raw combined data, which allocates knots spatially with a density proportional to sampling intensity. After the locations of the 200 knots had been determined, they were held fixed when model parameters were estimated.

To be able to calculate quantities for GOM red snapper, it was necessary to produce a two-dimensional spatial grid covering the spatial distribution range of the fish population, referred to as a "prediction grid," as described in Supplementary Appendix S2. Next, to construct indices for GOM red snapper with the fitted spatio-temporal models, we assumed that the Gaussian Markov random field in each cell of the prediction grid for GOM red snapper was equal to the value of the random field at the closest knot. Consequently, the surface area A_j (in km^2) associated with knot j was calculated as the number of cells of the prediction grid for GOM red snapper associated with knot j times the surface areas of these cells. It was then possible to estimate GOM red snapper biomass across the US GOM in year t, $\hat{B}(t)$, as:

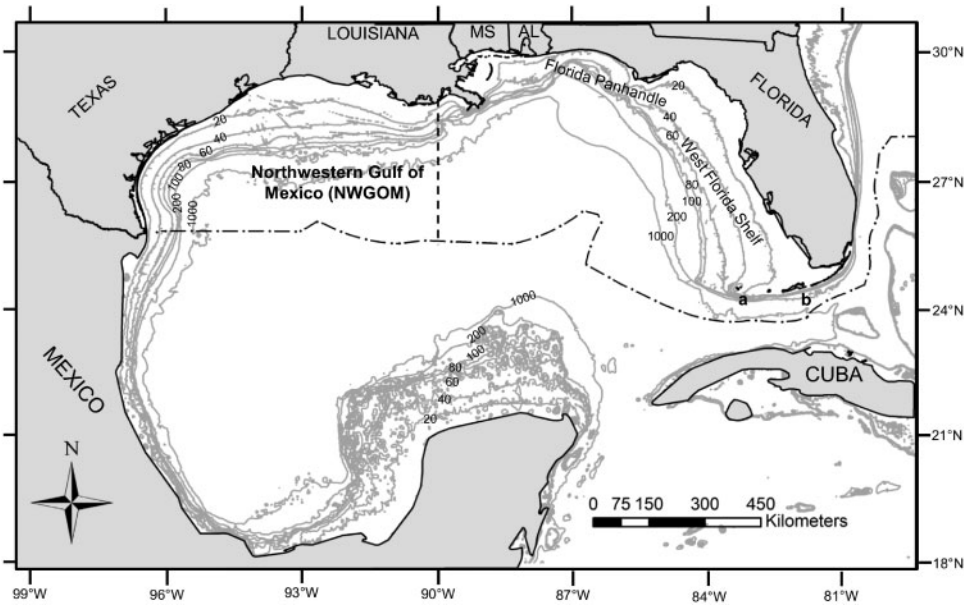

Figure 1. Map of the GOM. Important features are labelled and include: the NWGOM, the Florida Panhandle, the West Florida Shelf, the Dry Tortugas (a), and the Florida Keys (b). Depth contours are labelled in 20-, 40-, 60-, 80-, 100-, 200-, and 1000-m contours. MS, Mississippi; AL, Alabama. The black dashed-dotted line delineates the US exclusive economic zone.

$$\hat{B}(t) = \sum_{j=1}^{n_j} A_j \times \hat{n}(j,t) \times \hat{w}(j,t), \quad (7)$$

Ultimately, the indices for GOM red snapper were obtained by dividing the $\hat{B}(t)$ estimates by their mean value over the period 2006–2014. This definition of indices of relative biomass is very often used by fisheries scientists, as it facilitates a rapid interpretation of trends in relative biomass (Grüss, Walter, *et al.*, 2019).

The fitted spatio-temporal models were also employed to estimate eastward and northward COGs and effective area occupied. Eastward COG in year t, $\mathrm{ECOG}(t)$, is computed as follows (Thorson, Pinsky, *et al.*, 2016):

$$\mathrm{ECOG}(t) = \sum_{j=1}^{n_j} x_j \left\{ \frac{A_j \times \hat{n}(j,t) \times \hat{w}(j,t)}{\hat{B}(j,t)} \right\} \quad (8)$$

where x_j is the value of eastings (in km) in knot j. Northward COG in year t, $\mathrm{NCOG}(t)$, is calculated in a similar way, except that x_j is replaced with y_j, the value of northings (in km) in knot j, in Equation (8). Finally, the effective area occupied is estimated by dividing the estimated biomass [given by Equation (7)] by the average biomass-density. The average biomass-density in year t, $D(t)$, is computed as follows (Thorson, Pinsky, *et al.*, 2016):

$$D(t) = \sum_{j=1}^{n_j} \hat{n}(j,t) \times \hat{w}(j,t) \left\{ \frac{A_j \times \hat{n}(j,t) \times \hat{w}(j,t)}{\hat{B}(j,t)} \right\} \quad (9)$$

Simulation experiment

We also used the "bootstrap simulator" included in R package "VAST" (Thorson, 2019) to evaluate the accuracy, error, and confidence interval coverage of the indices predicted by the spatio-temporal model fitted to combined data, under "spatial" and "temporal" scenarios. Conditional on the MLE for fixed effects estimated for GOM red snapper, new random effects and new data were simulated, and it was then possible to fit the new data to the spatio-temporal model again and to compare the estimated indices to the "true" indices of relative biomass. Here, the spatial scenarios consisted of either collecting or not collecting biomass data in the NWGOM, while the temporal scenarios consisted of either collecting or not collecting biomass data in early years (i.e. over the period 2006–2008). Under the temporal scenario where biomass data were not collected in early years, it was necessary to constrain intercepts β_w to follow a random walk across years to allow intercepts to be estimated in years with no biomass-sampling data (Thorson, 2019). Under the spatial scenarios, we also compared the performance of spatio-temporal models fitted to combined vs. biomass-only data; this was not possible under the temporal scenarios, where biomass data were available for only 6 years (vs. 9 years for combined data).

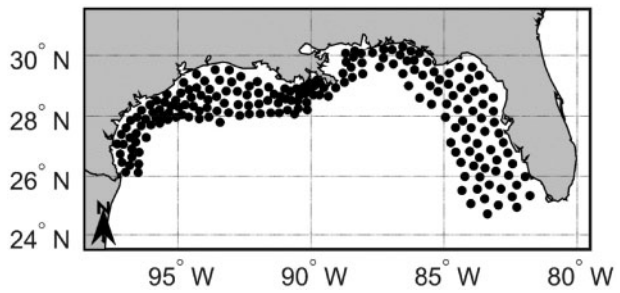

Figure 2. Spatial distribution of the "knots" used by the spatio-temporal models fitted to combined (i.e. encounter/non-encounter plus count plus biomass) data for GOM red snapper (*L. campechanus*). Knots serve to approximate all the spatial and spatio-temporal variation terms defined over the study spatial domain, such that spatio-temporal models track the value of these terms at each knot.

We considered 100 replicates for each scenario in this experiment. We evaluated three performance metrics, which are described in Supplementary Appendix S2: a metric of bias showing whether changes in the true index of relative biomass are accurately estimated (the closer to 1 the better; Thorson *et al.*, 2015); root mean squared error (the lower the better; Stow *et al.*, 2009); and coverage for a 50% confidence interval (henceforth "coverage"; the closer to 50% the better; Bolker, 2008).

Results

Demonstration for GOM red snapper

The TRAWL biomass dataset provided the largest number of data points (6358 samples across the 9 years). Combining BLL encounter/non-encounter and PELACTR count data with TRAWL biomass data increased the number of data points to 8687 (i.e. by ∼36.6%; Supplementary Figure S3). The spatial footprints of the TRAWL, BLL, and PELACTR programs were relatively similar (Figure 3a–h). However, the PELACTR program operated in deeper waters than the TRAWL program, while the BLL program sampled areas that were not monitored by the TRAWL program, including untrawlable sites off West Florida. The spatial distribution of GOM red snapper predicted by the models fitted to individual datasets (i.e. encounters/non-encounters, counts, or biomasses) were comparable (Figure 3i–k). These models all predicted red snapper to be mainly distributed on the shelves of Texas, Louisiana, Mississippi, and Alabama; to have a higher probability of encounter or higher abundance/biomass in the Florida Panhandle region than on the West Florida Shelf; and in its southernmost distribution areas (i.e. off Southwestern Florida), to be concentrated near the Dry Tortugas and the Florida Keys, at depths ranging between 20 and 60 m. The model fitted to combined data provided similar insights into red snapper spatial distribution patterns (Figure 3l and Supplementary Figure S4).

The models fitted to biomass-only or combined data predicted comparable indices, COGs and effective areas occupied (Figure 4a–d). Both models predicted no overall change in red snapper index over the period 2006–2014 (Figure 4a). However, TRAWL biomass data are also available for 2015, and models fitted to biomass-only or combined data for the period 2006–2015 predicted an overall increase in red snapper index over the period 2006–2015, consistent with the relative biomass estimates from the latest GOM red snapper assessment (Figure 4a). Moreover, the models fitted to biomass-only and combined data both predicted that red snapper COG moved both westward and southward over the period 2006–2014 (Figure 4b and c) and that red snapper distribution contracted from 2006 to 2011 (i.e. a decrease in effective area occupied) and then expanded from 2011 to 2014 (Figure 4d).

Using combined rather than biomass-only data reduced the log-standard error (log-SE) of the index and the SE of eastward COG and, overall, very slightly decreased the SE of northward COG and very slightly increased the log-SE of effective area occupied (Table 1 and Figure 4e–h). Moreover, maps of the log-SEs of red snapper biomass revealed that employing combining data reduced the log-SE of biomass in most of red snapper distribution areas, particularly off West Florida and in the deepest areas of the NWGOM shelf (Figure 5a and Supplementary Figure S5). On average over all years, using combined data decreased the log-SE of biomass in 65.3% of red snapper distribution areas.

Looking at the log-SEs of the indices predicted by models fitted to different combinations of encounter/non-encounter, count, and biomass data, it appears that the reduction of the log-SE of the index is due to the addition of encounter/non-encounter data to the dataset provided to VAST (Supplementary Figure S6). The log-SE of the index is greater when VAST is provided with count data or a combination of count and biomass data than when VAST is provided with a combination of encounter/non-encounter and biomass data or combined data. We then examined the fixed parameters estimated by the spatio-temporal models and their SEs. This revealed that the increase in the log-SE of the index due to count data stems from large changes in the year effects β_n and an important increase in the SE of these year effects (Supplementary Figure S7).

We now examine a "spatial" scenario, where all the biomass data except for 98% of those for the NWGOM are considered. Under the spatial scenario, the models fitted to biomass-only or combined data keep providing similar insights into the spatial distribution patterns of red snapper (Supplementary Figure S8). The decrease in the log-SE of the index and the SEs of the eastward and northward COGs, as well as the increase in the log-SE of effective area occupied, due to the substitution of biomass-only with combined data, are much more pronounced under the spatial than under the base scenario; this is particularly true for the SEs of the eastward and northward COGs (Table 1 and Figure 6a–d). Moreover, using combined data still leads to a reduction in the log-SE of biomass in most of red snapper distribution areas (Figure 5b and Supplementary Figure S9). This reduction is more pronounced under the spatial than under the base scenario (occurs in 73.7% of red snapper distribution areas on average over all years under the spatial scenario vs. 65.3% under the base scenario).

Next, we examine a "temporal" scenario, where all the biomass data except for 98% of those for the period 2006–2008 are considered. Under the temporal scenario, the models fitted to biomass-only or combined data keep providing similar insights into the spatial distribution patterns of red snapper (Supplementary Figure S10). The decrease in the log-SE of the index and the SEs of the eastward and northward COGs due to the substitution of biomass-only with combined data are less pronounced under the temporal than under the base scenario (Table 1 and Figure 6e–g). However, the increase in the log-SE of effective area occupied due to the use of combined data in lieu of biomass-only data is slightly more pronounced under the temporal than under the base scenario (Table 1 and Figure 6h). Finally, employing combined data still leads to a reduction in the log-SE of biomass in most of red snapper distribution areas (Figure 5c and Supplementary Figure S11). This reduction is much more pronounced under the temporal than under the base scenario (occurs in 90.1% of red snapper distribution areas on average over all years under the temporal scenario vs. 65.3% under the base scenario).

Simulation experiment

We also explored spatial and temporal scenarios within a simulation experiment, where biomass data were, respectively, either collected or not collected in the NWGOM and either collected or not collected in early years. Depending on the replicates considered, the model fitted to combined data produced indices that matched the "true" indices very well (Figure 7a–d) or not

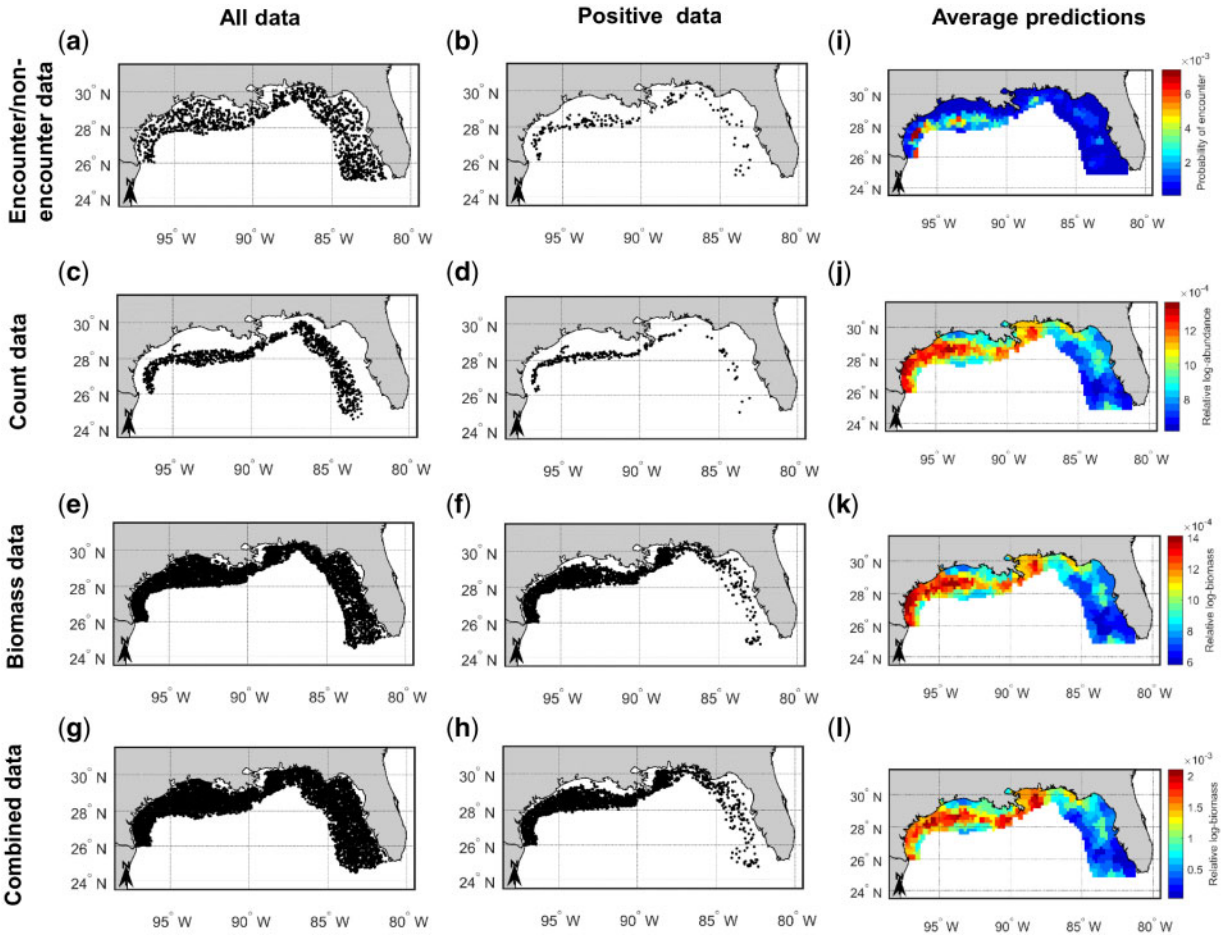

Figure 3. Spatial distribution of the (a, b) encounter/non-encounter, (c, d) count, (e, f) biomass, and (g, h) combined (i.e. encounter/non-encounter plus count plus biomass) data for GOM red snapper (*L. campechanus*), and (i–l) average spatial predictions over the period 2006–2014 made with spatio-temporal models fitted to these data (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.).

(Figure 7e–h). Not collecting biomass data in the NWGOM or in early years generally resulted in more negatively biased and more imprecise indices; results in terms of bias were more pronounced under the temporal than under the spatial scenario, while the opposite was true for results in terms of error (Figure 8a–d). Under both the temporal and spatial scenarios, models fitted to combined data had reasonable coverage (Figure 8e and f). Under the spatial scenarios, we also conducted a comparison of the performance of spatio-temporal models fitted to combined vs. biomass-only data. This comparison revealed that the indices produced by models fitted to biomass-only data are often less biased than those produced by models fitted to combined data, although the models fitted to biomass-only and combined data have comparable error and coverage (Figure 9).

Discussion

This study confirmed earlier IPM studies that combining multiple data types (i.e. using data-integrated models) is overall beneficial to ecological investigations. A diversity of data types, including encounters/non-encounters, counts, biomasses, and others, are collected in the marine environment (Grüss, Perryman, *et al.*, 2018), and combining these different data types is straightforward and can be achieved if: (i) the different datasets considered were collected separately; (ii) the likelihoods for individual datasets have parameters in common; and (iii) the individuals included in these different datasets can be assumed to belong to the same population. The three aforementioned conditions will very often be fulfilled with "unmarked" data types such as surveys (Schaub and Abadi, 2011; Zipkin *et al.*, 2017).

The spatio-temporal model fitted to combined data provided insights into red snapper spatial distribution patterns in the US GOM, which were similar to the insights from an earlier spatio-temporal model fitted to multiple encounter/non-encounter datasets (Grüss, Perryman, *et al.*, 2018) and a generalized linear model fitted to multiple count datasets (Karnauskas *et al.*, 2017). However, relying on biomass and count data in addition to encounter/non-encounter data in this study allowed us to provide more information for population and habitat assessments. First, we were able to reconstruct indices of relative biomass for GOM red snapper population assessments. Then, we produced information on red snapper COGs and effective area occupied for habitat assessments. We found that, over the study period (2006–2014), the COG of red snapper moved both westward and southward, while its effective area occupied decreased between 2006 and 2011 and then increased between 2011 and 2014. To our knowledge, this was the first time that patterns of distribution shifts and

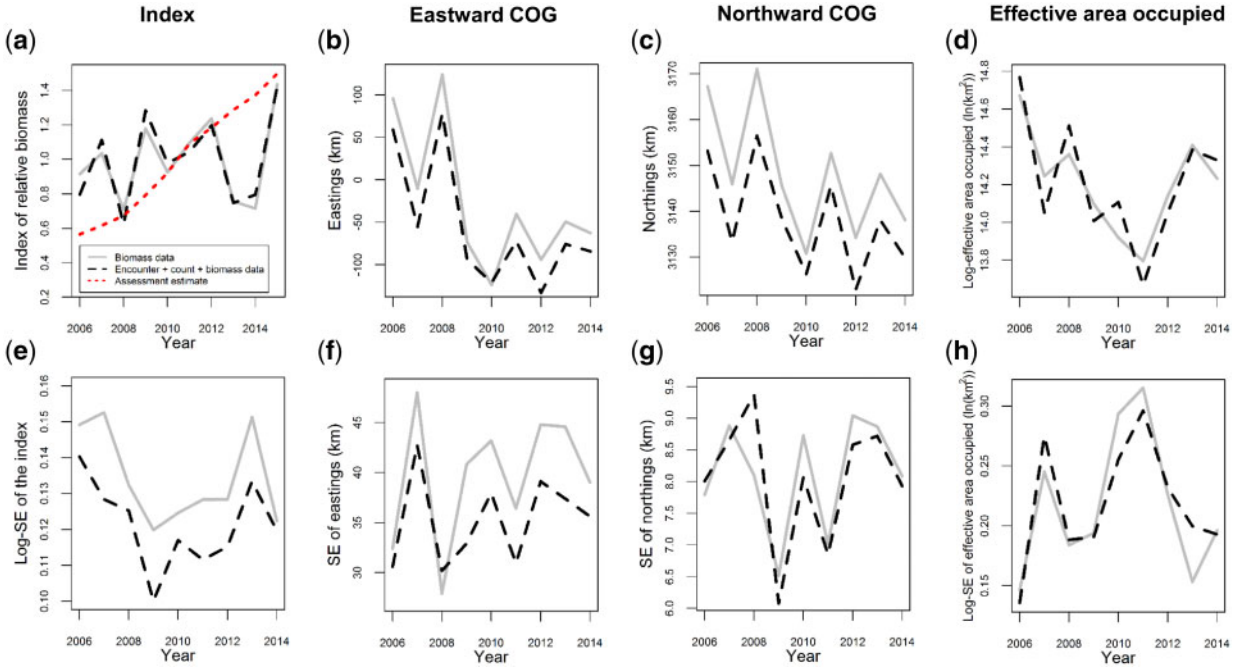

Figure 4. Predictions of the spatio-temporal models fitted to biomass-only data (full grey lines) or combined (i.e. encounter/non-encounter plus count plus biomass) data (dashed black lines) for GOM red snapper (*L. campechanus*). (a) Indices of relative biomass and (e) their log-SE. (b, c) COGs of red snapper (km) and (f, g) their SE (km). (d) Log-effective areas occupied by red snapper [ln(km^2)] and (h) their SE [ln(km^2)]. In (a), the relative biomass estimate from the latest GOM red snapper population assessment (SEDAR 52, 2018) is also provided (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.).

range expansion/contraction were examined for GOM red snapper, despite the strong socio-economic importance of this fish population. It is important to note that changes in COGs can be a product of no actual movement, but can rather be due to different spatial patterns of mortality and recruitment; regardless, the westward and southward shift of GOM red snapper COG between 2006 and 2014 reflect biomass trajectories in the eastern US GOM (flat and decreasing) vs. western US GOM (sharply increasing) estimated in the latest GOM red snapper population assessment (SEDAR 52, 2018).

However, while our study delivered useful information for GOM red snapper population assessments, we recommend future studies to consider a longer time period. In this study, we used 9 years of data collected by three fisheries-independent programs (the TRAWL biomass, the PELACTR count, and the BLL encounter/non-encounter datasets) that can be shared publicly (https://figshare.com/articles/Red_snapper_dataset/7451276). Yet, a large number of fisheries-independent and fisheries-dependent datasets (of which many are confidential) can provide biomass, count, and encounter/non-encounter data, which can be employed to reconstruct patterns of spatial distribution, distribution shifts, range expansion/contraction, and relative biomass for GOM red snapper since the early 1980s (Grüss, Perryman, *et al.*, 2018). This endeavour is desirable, given that changes in GOM red snapper biomass were more pronounced from the early 1980 to the present than over the short time period considered in this study (2006–2014) (SEDAR 52, 2018).

When working with the original dataset for GOM red snapper (i.e. under the base scenario), using combined rather than biomass-only data decreased the SEs of the index and the COGs, as well as the log-SE of biomass in most of red snapper distribution areas, particularly in those areas poorly or not sampled by the TRAWL biomass dataset. Substituting biomass-only data with combined data also led to an increase in the log-SE of effective area occupied, although this increase was very small (+2%). Thus, overall, this study concurs with IPM studies, which showed that combining data resulted in more precise estimates of demographic quantities (Brooks *et al.*, 2004; Schaub *et al.*, 2007; Abadi *et al.*, 2012; Dorazio, 2014; Zipkin *et al.*, 2017). However, we were expecting a larger reduction in the log-SE of the index than the one found in this study (−9.7%). This reduction was due to the combination of biomass and BLL encounter/non-encounter data. The PELACTR count data did not contribute to decreasing the log-SE of the index, because they greatly altered the year effects on numbers-density (β_n parameters) and greatly reduced the precision of these year effects. Therefore, we conclude that data-integrated models will not inevitably result in substantially increased precision, and we recommend that future studies explore results for each candidate encounter/non-encounter and count dataset individually, as a preliminary diagnostic to identify which dataset could contribute to noticeable changes in estimated precision.

We chose to analyse data for GOM red snapper for the period 2006–2014 to ensure that all data types were available across the entire spatial domain for all years. In the real world, we suspect that many monitoring programs collecting the different data types may not cover or may poorly cover some subregions of the geographic area of interest. These scenarios reflect the situation of, e.g. some snapper and grouper populations such as the GOM populations of cubera snapper (*Lutjanus cyanopterus*) and goliath grouper (*Epinephelus itajara*) (Grüss, Biggs, *et al.*, 2018). Moreover, for many fish and invertebrate populations and life

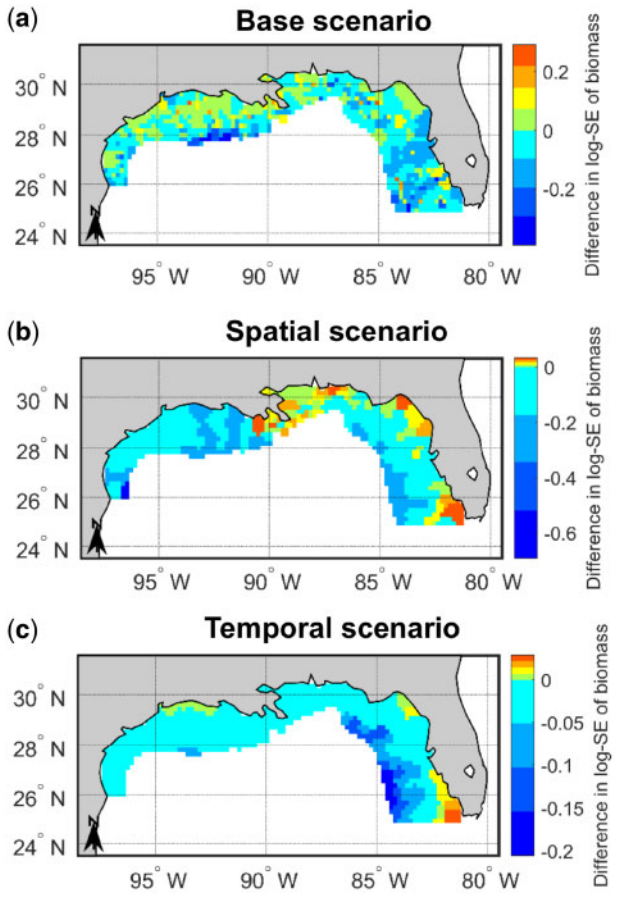

Figure 5. Comparison of the average log-SEs of GOM red snapper (*L. campechanus*) biomass over the period 2006–2014 predicted by the spatio-temporal model fitted to biomass-only data and those predicted by the spatio-temporal model fitted to combined (i.e. encounter/non-encounter plus count plus biomass) data, under the (a) "base," (b) "spatial," and (c) "temporal" scenarios. A positive (negative) difference indicates that using combined data in lieu of biomass-only data increased (decreased) the log-SE of red snapper biomass locally. In the base scenario, all biomass data were considered. In the spatial scenario, all the biomass data except for 98% of those for the NWGOM were considered. In the temporal scenario, all the biomass data except for 98% of those for the period 2006–2008 were considered (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.).

stages, biomass data are unavailable for early years in a time series, while other datasets (e.g. count datasets) cover these early years. Thus, we also examined spatial and temporal scenarios, where red snapper biomass data were very poorly or not sampled in an entire subregion or in early years. Under the spatial scenarios, employing combined rather than biomass-only data substantially reduced the SEs of the index and the COGs, while still increasing the log-SE of effective area occupied. Under the temporal scenarios, using combined rather than biomass-only data offered reduced benefits compared to the spatial scenarios and, to a lesser extent, the base scenario. As expected, when no biomass data were collected in an entire subregion, the spatio-temporal model fitted to combined data predicted more biased and more imprecise indices. Under the temporal scenarios, the spatio-temporal model fitted to combined data resulted in more biased, but more precise, indices than under the spatial scenarios. Yet, under both the spatial and temporal scenarios, a spatio-temporal model fitted to combined data also had reasonable coverage, i.e. confidence intervals that were not too narrow nor too wide and, thus, represented uncertainty adequately.

We envision three avenues for future research. First, future studies should consider additional data types, particularly those collected by opportunistic surveys that are the primary source of data for many populations, such as binned counts (e.g. collected for GOM goliath grouper by the Reef Environmental Education Foundation Fish Survey; Thorson *et al.*, 2014) or maximum school sizes (e.g. collected for Indo-Pacific elasmobranchs by recreational scuba divers; Ward-Paige and Lotze, 2011). Presence-only data are also the main source of data for many populations (e.g. habitat-forming invertebrate populations; Vierod *et al.*, 2014; Guinotte *et al.*, 2017). However, working with presence-only data is running the risk of mistakenly identifying areas that are well monitored as high-density areas (Fithian *et al.*, 2015). Therefore, we recommend future studies using our modelling framework and presence-only data to: (i) generate pseudo-absences (e.g. by recovering transect information or by selecting at random within the region covered by the presence data at least one order of magnitude more pseudo-absences than presences; Grüss, Drexler, *et al.*, 2019) so as to work with presence/pseudo-absence data; and (ii) expand our data-integrated model in the case of presence/pseudo-absence data so as to account for monitoring intensity and the covariates influencing this process (Fithian *et al.*, 2015). Second, our modelling framework should also be employed to support climate-vulnerability and ecosystem assessments. For example, relevant environmental covariates could be introduced in the state-process equations [Equation (5)], and the model fitted to combined data could then be used to forecast spatial distribution patterns and population trends in

Table 1. Changes (in %) in some of the SEs predicted by spatio-temporal models for GOM red snapper (*L. campechanus*) under the "base," "spatial," and "temporal" scenarios when these models are fitted using encounter/non-encounter, count, and biomass data instead of just biomass data.

	Base scenario, %	Spatial scenario, %	Temporal scenario, %
Change in the log-SE of the index of relative biomass	−9.7	−12.9	−6.3
Change in the SE of the eastward COG	−10.3	−29.8	−8.1
Change in the SE of the northward COG	−1.1	−26.6	−0.9
Change in the log-SE of effective area occupied	+2.0	+9.6	+4.5

In the base scenario, all biomass data were considered. In the spatial scenario, all the biomass data except for 98% of those for the NWGOM were considered. In the temporal scenario, all the biomass data except for 98% of those for the period 2006–2008 were considered.

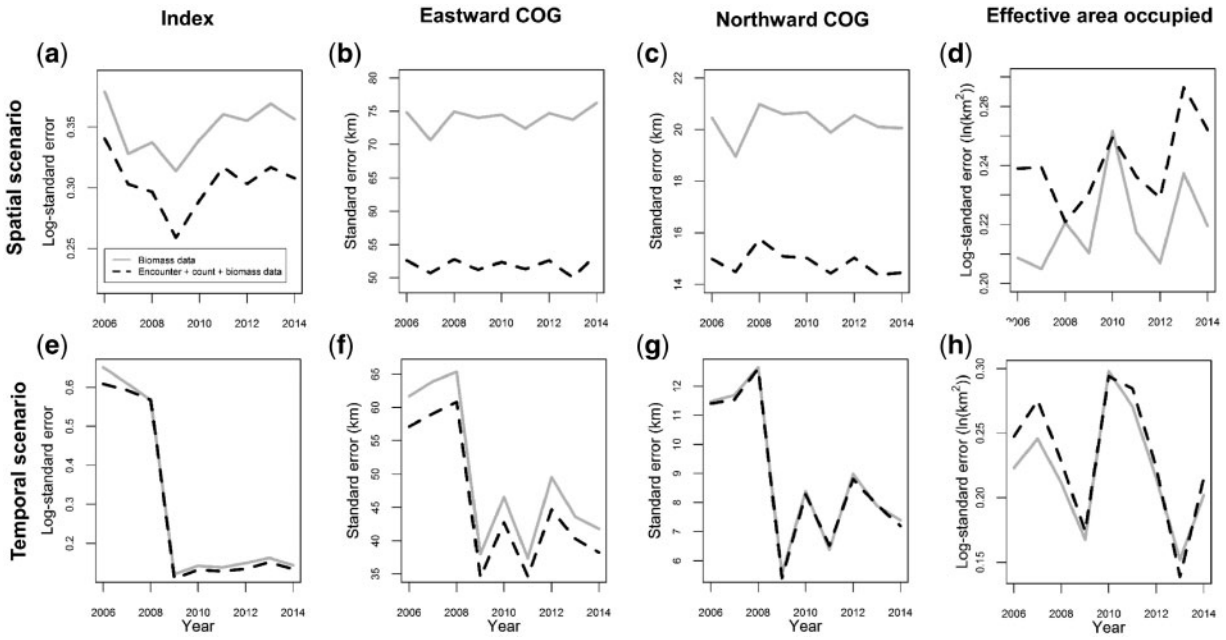

Figure 6. Predictions of the spatio-temporal models fitted to biomass-only data (full grey lines) or combined (i.e. encounter/non-encounter plus count plus biomass) data (dashed black lines) for GOM red snapper (*L. campechanus*), under the (a–d) "spatial" and (e–h) "temporal" scenarios. (a, e) Log-SE of the indices of relative biomass. (b, c and f, g) SE of the COGs of red snapper (km). (d, h) Log-SE of the effective area occupied by red snapper [ln(km^2)]. In the spatial scenario, all the biomass data except for 98% of those for the NWGOM were considered. In the temporal scenario, all the biomass data except for 98% of those for the period 2006–2008 were considered.

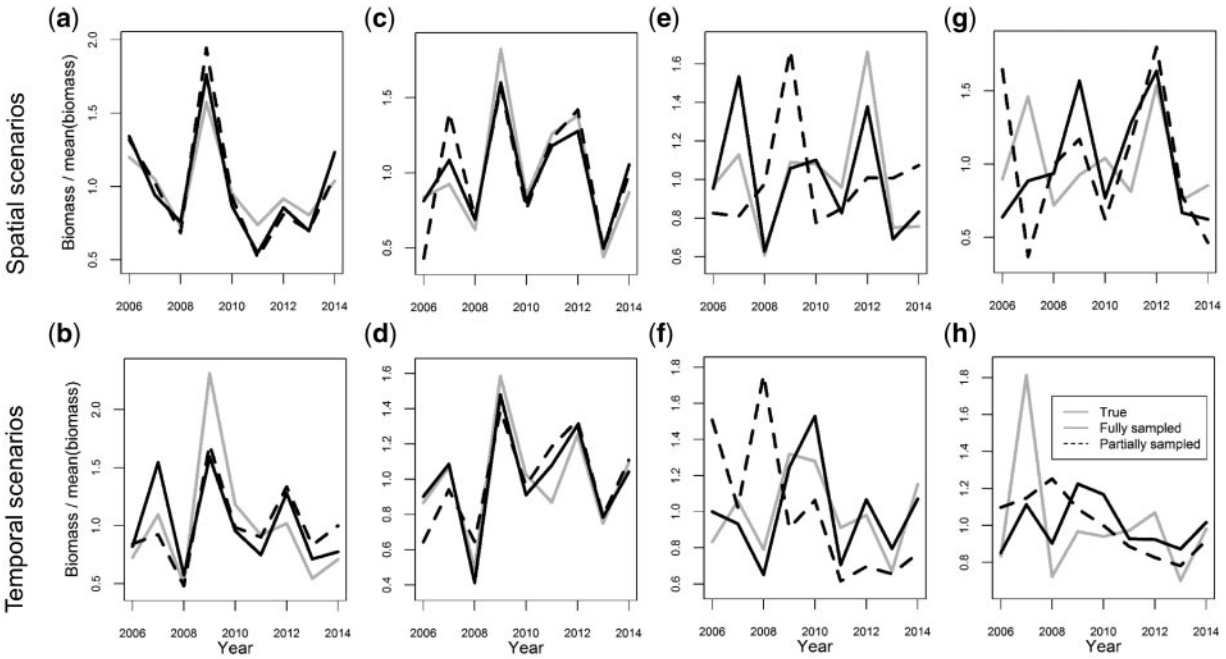

Figure 7. Some of the indices of relative biomass estimated by spatio-temporal models fitted to combined (i.e. encounter/non-encounter plus count plus biomass) data under the (a, c, e, g) "spatial" and (b, d, f, h) "temporal" scenarios tested within the simulation experiment. In the spatial scenarios, the "fully sampled population" is a population for which biomass data are collected all over the US GOM, while the "partially sampled population" is a population for which biomass data are not collected in the NWGOM. In the temporal scenarios, the "fully sampled population" is a population for which biomass data are collected in all years, while the "partially sampled population" is a population for which biomass data are not collected over the period 2006–2008.

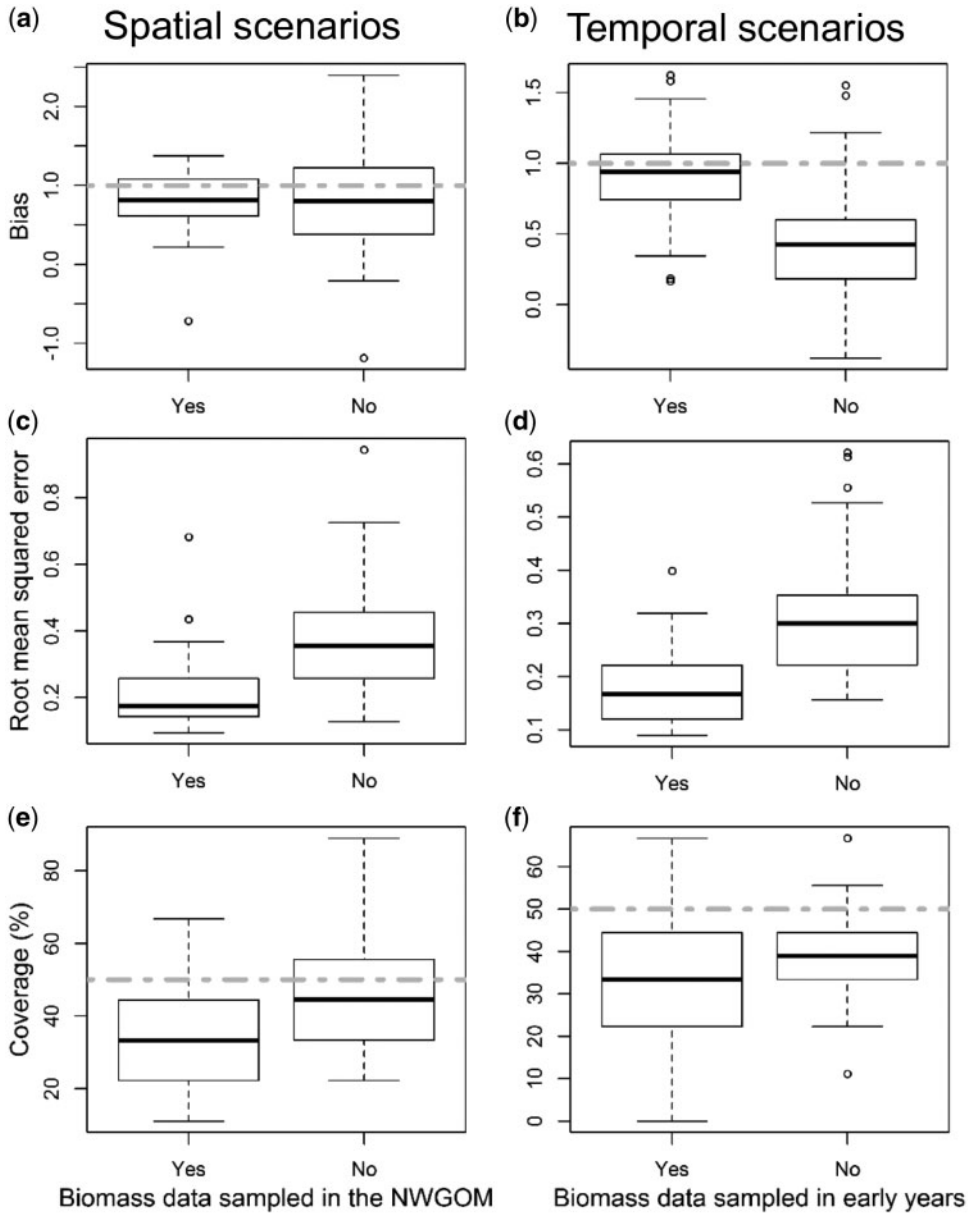

Figure 8. (a, b) Bias (the closer to 1 the better), (c, d) root mean squared error (the lower the better), and (e, f) coverage (in %; the closer to 50% the better) of the indices of relative biomass estimated by spatio-temporal models fitted to combined (i.e. encounter/non-encounter plus count plus biomass) data under the (a, c, e) "spatial" and (b, d, f) "temporal" scenarios tested within the simulation experiment. In the spatial scenarios, biomass data were either collected or not collected in the NWGOM. In the temporal scenarios, biomass data were either collected or not collected in early years (i.e. over the period 2006–2008).

response to climate change (Thorson, 2018b). Third, we recommend that future research utilizes our modelling framework within a simulation experiment for monitoring program optimization, whereby a spatio-temporal model is fitted to combined data to produce simulated sampling data under alternative sampling designs. Each simulated dataset would then be tested, and the average performance of each sampling design would be evaluated to ultimately identify sampling designs that provide an optimal utilization of available resources (Reich *et al.*, 2018; Thorson, 2019).

In conclusion, the ability to fit spatio-temporals model to different data types is valuable for assisting the diverse assessments that fisheries scientists are tasked carrying out, and we recommend future studies to further explore the performance of our modelling framework, particularly for populations and life stages lacking biomass data for entire subregions and/or time periods. Importantly, our modelling framework may, at long last, allow for the reconstruction of population trends for numerous data-poor, yet socio-economically important populations and, consequently, for more elaborated stock assessments and management measures for these populations (e.g. many snapper and grouper populations of the US GOM; Grüss, Biggs, *et al.*, 2018; Grüss, Perryman, *et al.*, 2018). Moreover, the maps of relative biomass produced with our modelling framework for critical life stages

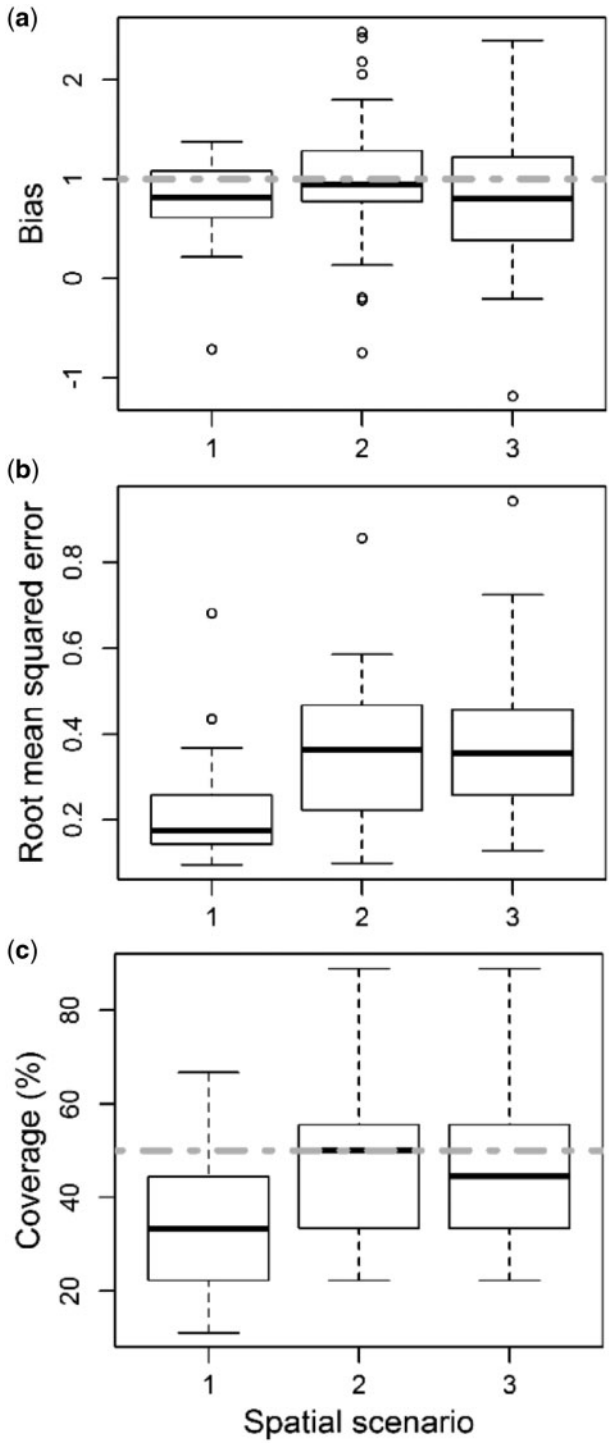

Figure 9. (a) Bias (the closer to 1 the better), (b) root mean squared error (the lower the better), and (c) coverage (in %; the closer to 50% the better) of the indices of relative biomass estimated by spatio-temporal models under the "spatial" scenarios tested within the simulation experiment. In Scenario 1, a spatio-temporal model was fitted to combined (i.e. encounter/non-encounter plus count plus biomass) data, which included biomass data collected all over the US GOM (including the NWGOM). In Scenario 2, a spatio-temporal model was fitted to biomass-only data collected only in the northeastern GOM. Finally, in Scenario 3, a spatio-temporal model was fitted to combined data, which included biomass data collected only in the northeastern GOM.

(e.g. young-of-the-year, spawners) will provide robust insights into the location of biomass hotspots for these critical life stages and, subsequently, will enable more effective protection of essential fish habitat (Laman *et al.*, 2018).

Supplementary data

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

Acknowledgements

We thank two NOAA internal reviewers (Elizabeth Babcock and John Walter), as well as two anonymous reviewers, whose comments have improved the quality of our manuscript. We are grateful to Jeff Rester and Adam Pollack for having provided monitoring data for this study. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect those of NOAA or the Department of Commerce.

Author contributions

AG and JTT conceived the ideas and designed the methods, analyzed the results, and wrote the manuscript.

References

Abadi, F., Gimenez, O., Jakober, H., Stauber, W., Arlettaz, R., and Schaub, M. 2012. Estimating the strength of density dependence in the presence of observation errors using integrated population models. Ecological Modelling, 242: 1–9.

Berg, C. W., Nielsen, A., and Kristensen, K. 2014. Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models. Fisheries Research, 151: 91–99.

Bolker, B. M. 2008. Ecological Models and Data in R. Princeton University Press, Princeton, NJ.

Brooks, S. P., King, R., and Morgan, B. J. T. 2004. A Bayesian approach to combining animal abundance and demographic data. Animal Biodiversity and Conservation, 27: 515–529.

Cao, J., Thorson, J. T., Richards, R. A., and Chen, Y. 2017. Spatiotemporal index standardization improves the stock assessment of northern shrimp in the Gulf of Maine. Canadian Journal of Fisheries and Aquatic Sciences, 74: 1781–1793.

Dail, D., and Madsen, L. 2011. Models for estimating abundance from repeated counts of an open metapopulation. Biometrics, 67: 577–587.

Dolder, P. J., Thorson, J. T., and Minto, C. 2018. Spatial separation of catches in highly mixed fisheries. Scientific Reports, 8: 13886.

Dorazio, R. M. 2014. Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. Global Ecology and Biogeography, 23: 1472–1484.

Federal Register. 2008. Magnuson-Stevens Act Provisions; Annual Catch Limits; National Standard Guidelines. Proposed Rule, Vol. 73: No. 111/Monday, June 9, 2008/Proposed Rules.

Fithian, W., Elith, J., Hastie, T., and Keith, D. A. 2015. Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods in Ecology and Evolution, 6: 424–438.

Grüss, A., Biggs, C., Heyman, W. D., and Erisman, B. 2018. Prioritizing monitoring and conservation efforts for fish spawning aggregations in the US Gulf of Mexico. Scientific Reports, 8: 8473.

Grüss, A., Drexler, M., and Ainsworth, C. H. 2014. Using delta generalized additive models to produce distribution maps for spatially explicit ecosystem models. Fisheries Research, 159: 11–24.

Grüss, A., Drexler, M. D., Ainsworth, C. H., Babcock, E. A., Tarnecki, J. H., and Love, M. S. 2018. Producing distribution maps for a

spatially-explicit ecosystem model using large monitoring and environmental databases and a combination of interpolation and extrapolation. Frontiers in Marine Science, 5: 16.

Grüss, A., Drexler, M. D., Chancellor, E., Ainsworth, C. H., Gleason, J. S., Tirpak, J. M., Love, M. S. *et al.* 2019. Representing species distributions in spatially-explicit ecosystem models from presence-only data. Fisheries Research, 210: 89–105.

Grüss, A., Perryman, H. A., Babcock, E. A., Sagarese, S. R., Thorson, J. T., Ainsworth, C. H., Anderson, E. J. *et al.* 2018. Monitoring programs of the US Gulf of Mexico: inventory, development and use of a large monitoring database to map fish and invertebrate spatial distributions. Reviews in Fish Biology and Fisheries, 28: 667–691.

Grüss, A., Thorson, J. T., Babcock, E. A., and Tarnecki, J. H. 2018. Producing distribution maps for informing ecosystem-based fisheries management using a comprehensive survey database and spatio-temporal models. ICES Journal of Marine Science, 75: 158–177.

Grüss, A., Thorson, J. T., Sagarese, S. R., Babcock, E. A., Karnauskas, M., Walter, J. F. III and Drexler, M. 2017. Ontogenetic spatial distributions of red grouper (*Epinephelus morio*) and gag grouper (*Mycteroperca microlepis*) in the US Gulf of Mexico. Fisheries Research, 193: 129–142.

Grüss, A., Walter, J. F. III, Babcock, E. A., Forrestal, F. C., Thorson, J. T., Lauretta, M. V., and Schirripa, M. J. 2019. Evaluation of the impacts of different treatments of spatio-temporal variation in catch-per-unit-effort standardization models. Fisheries Research, 213: 75–93.

Grüss, A., Yemane, D., and Fairweather, T. P. 2016. Exploring the spatial distribution patterns of South African Cape hakes using generalised additive models. African Journal of Marine Science, 38: 395–409.

Guinotte, J. M., Georgian, S., Kinlan, B. P., Poti, M., and Davies, A. J. 2017. Predictive habitat modeling for deep-sea corals in U.S. waters. *In* The State of Deep-sea Coral and Sponge Ecosystems of the United States. NOAA Technical Memorandum NMFS-OHC-4. Ed. by T. F. Hourigan, P. J. Etnoyer, and S. D. Cairns. National Marine Fisheries Service, Silver Spring, MD.

Henwood, T., Ingram, W., and Grace, M. 2006. Shark/snapper/grouper longline surveys. Southeast Data Assessment and Review (SEDAR), SEDAR7-DW-08, North Charleston, SC. 22 pp.

Kai, M., Thorson, J. T., Piner, K. R., and Maunder, M. N. 2017. Spatiotemporal variation in size-structured populations using fishery data: an application to shortfin mako (*Isurus oxyrinchus*) in the Pacific Ocean. Canadian Journal of Fisheries and Aquatic Sciences, 74: 1765–1780.

Karnauskas, M., Walter, J. F. III, Campbell, M. D., Pollack, A. G., Drymon, J. M., and Powers, S. 2017. Red Snapper distribution on natural habitats and artificial structures in the northern Gulf of Mexico. Marine and Coastal Fisheries, 9: 50–67.

Kass, R. E., and Steffey, D. 1989. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). Journal of the American Statistical Association, 84: 717–726.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. 2016. TMB: automatic differentiation and Laplace approximation. Journal of Statistical Software, 70: 1–20.

Laman, E. A., Rooper, C. N., Turner, K., Rooney, S., Cooper, D. W., and Zimmermann, M. 2018. Using species distribution models to describe essential fish habitat in Alaska. Canadian Journal of Fisheries and Aquatic Sciences, 75: 1230–1255.

Lebreton, J.-D., Morgan, B. J., Pradel, R., and Freeman, S. N. 1995. A simultaneous survival rate analysis of dead recovery and live recapture data. Biometrics, 51: 1418–1428.

Maunder, M. N. 2004. Population viability analysis based on combining Bayesian, integrated, and hierarchical analyses. Acta Oecologica, 26: 85–94.

MSRA (Magnuson-Stevens Fishery Conservation and Management Reauthorization Act). 2006. U.S. Public Law 109–479, 120 Statute 3575.

National Marine Fisheries Service (NMFS). 2017. Fisheries economics of the United States 2015: economics and sociocultural status and trends series. NOAA Technical Memorandum. NOAA, Department of Commerce, Silver Spring, MD. 245 pp.

Pauly, D. 1995. Anecdotes and the shifting baseline syndrome of fisheries. Trends in Ecology & Evolution, 10: 430–430.

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.

Pirtle, J. L., Shotwell, S. K., Zimmermann, M., Reid, J. A., and Golden, N. 2017. Habitat suitability models for groundfish in the Gulf of Alaska. Deep Sea Research Part II: Topical Studies in Oceanography, doi: 10.1016/j.dsr2.2017.12.005

Pollack, A. G., and Ingram, G. W. 2014. Smoothhound abundance indices from NMFS small pelagics surveys in the northern Gulf of Mexico. Southeast Data and Assessment and Review (SEDAR), SEDAR39-DW-08, North Charleston, SC. 14 pp.

Reich, B. J., Pacifici, K., and Stallings, J. W. 2018. Integrating auxiliary data in optimal spatial design for species distribution modelling. Methods in Ecology and Evolution, 9: 1626–1637.

Rester, J. K. 2017. SEAMAP Environmental and Biological Atlas of the Gulf of Mexico, 2015. Number 263. Gulf States Marine Fisheries Commission. 72 pp.

Rosenberg, A., Bigford, T. E., Leathery, S., Hill, R. L., and Bickers, K. 2000. Ecosystem approaches to fishery management through essential fish habitat. Bulletin of Marine Science, 66: 535–542.

Rossman, S., Yackulic, C. B., Saunders, S. P., Reid, J., Davis, R., and Zipkin, E. F. 2016. Dynamic N-occupancy models: estimating demographic rates and local abundance from detection-nondetection data. Ecology, 97: 3300–3307.

Runnebaum, J., Guan, L., Cao, J., O'Brien, L., and Chen, Y. 2018. Habitat suitability modeling based on a spatiotemporal model: an example for cusk in the Gulf of Maine. Canadian Journal of Fisheries and Aquatic Sciences, 75: 1–14.

Schaub, M., and Abadi, F. 2011. Integrated population models: a novel analysis framework for deeper insights into population dynamics. Journal of Ornithology, 152: 227–237.

Schaub, M., Gimenez, O., Sierro, A., and Arlettaz, R. 2007. Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. Conservation Biology, 21: 945–955.

Schnute, J. T. 1994. A general framework for developing sequential fisheries models. Canadian Journal of Fisheries and Aquatic Sciences, 51: 1676–1688.

SEDAR 52. 2018. SEDAR 52 Stock Assessment Report. Gulf of Mexico Red Snapper. http://www.sefsc.noaa.gov/sedar/.

Shelton, A. O., Thorson, J. T., Ward, E. J., and Feist, B. E. 2014. Spatial semiparametric models improve estimates of species abundance and distribution. Canadian Journal of Fisheries and Aquatic Sciences, 71: 1655–1666.

Stow, C. A., Jolliff, J., McGillicuddy, D. J. Jr, Doney, S. C., Allen, J. I., Friedrichs, M. A., Rose, K. A. *et al.* 2009. Skill assessment for coupled biological/physical models of marine systems. Journal of Marine Systems, 76: 4–15.

Thorson, J. T. 2018a. Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. Canadian Journal of Fisheries and Aquatic Sciences, 75: 1369–1382.

Thorson, J. T. 2018b. Forecast skill for predicting distribution shifts: a retrospective experiment for marine fishes in the Eastern Bering Sea. Fish and Fisheries, doi: 10.1111/faf.12330

Thorson, J. T. 2019. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fisheries Research, 210: 143–161.

Thorson, J. T., Fonner, R., Haltuch, M. A., Ono, K., and Winker, H. 2017. Accounting for spatiotemporal variation and fisher targeting when estimating abundance from multispecies fishery data. Canadian Journal of Fisheries and Aquatic Sciences, 74: 1794–1807.

Thorson, J. T., and Haltuch, M. A. 2018. Spatio-temporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. Canadian Journal of Fisheries and Aquatic Sciences, doi: 10.1139/cjfas-2018-0015

Thorson, J. T., and Kristensen, K. 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. Fisheries Research, 175: 66–74.

Thorson, J. T., Pinsky, M. L., and Ward, E. J. 2016. Model-based inference for estimating shifts in species distribution, area occupied and centre of gravity. Methods in Ecology and Evolution, 7: 990–1002.

Thorson, J. T., Rindorf, A., Gao, J., Hanselman, D. H., and Winker, H. 2016. Density-dependent changes in effective area occupied for sea-bottom-associated marine fishes. Proceedings of the Royal Society B, 283: 20161853.

Thorson, J. T., Scheuerell, M. D., Semmens, B. X., and Pattengill-Semmens, C. V. 2014. Demographic modeling of citizen science data informs habitat preferences and population dynamics of recovering fishes. Ecology, 95: 3251–3258.

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

Trenkel, V. M., Elston, D. A., and Buckland, S. T. 2000. Fitting population dynamics models to count and cull data using sequential importance sampling. Journal of the American Statistical Association, 95: 363–374.

Vierod, A. D., Guinotte, J. M., and Davies, A. J. 2014. Predicting the distribution of vulnerable marine ecosystems in the deep sea using presence-background models. Deep Sea Research Part II: Topical Studies in Oceanography, 99: 6–18.

Ward-Paige, C. A., and Lotze, H. K. 2011. Assessing the value of recreational divers for censusing elasmobranchs. PLoS One, 6: e25609.

Wilson, S., Gil-Weir, K. C., Clark, R. G., Robertson, G. J., and Bidwell, M. T. 2016. Integrated population modeling to assess demographic variation and contributions to population growth for endangered whooping cranes. Biological Conservation, 197: 1–7.

Zipkin, E. F., Rossman, S., Yackulic, C. B., Wiens, J. D., Thorson, J. T., Davis, R. J., and Grant, E. H. C. 2017. Integrating count and detection–nondetection data to model population dynamics. Ecology, 98: 1640–1650.

Handling editor: Ernesto Jardim