Spatiotemporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment

James T. Thorson and Melissa A. Haltuch

Abstract: Stock assessment models are fitted to abundance-index, fishery catch, and age–length–sex composition data that are estimated from survey and fishery records. Research has developed spatiotemporal methods to estimate abundance indices, but there is little research regarding model-based methods to generate age–length–sex composition data. We demonstrate a spatiotemporal approach to generate composition data and a multinomial sample size that approximates the estimated imprecision. A simulation experiment comparing spatiotemporal and design-based methods demonstrates a 32% increase in input sample size for the spatiotemporal estimator. A Stock Synthesis assessment used to manage lingcod (*Ophiodon elongatus*) in the California Current also shows a 17% increase in sample size and better model fit using the spatiotemporal estimator, resulting in smaller standard errors when estimating spawning biomass. We conclude that spatiotemporal approaches are feasible for estimating both abundance-index and compositional data, thereby providing a unified approach for generating inputs for stock assessments. We hypothesize that spatiotemporal methods will improve statistical efficiency for composition data in many stock assessments and recommend that future research explore the impact of including additional habitat or sampling covariates.

Résumé : Les modèles d'évaluation des stocks sont calés sur des données d'indice d'abondance, de prises des pêches et de composition selon l'âge, la longueur ou le sexe qui sont estimées à partir de registres de relevés et de pêche. Des travaux de recherche ont mené à l'établissement de méthodes spatiotemporelles pour estimer les indices d'abondance, mais peu de travaux ont touché aux méthodes basées sur les modèles pour produire des données de compositions selon l'âge, la longueur ou le sexe. Nous faisons la démonstration d'une approche spatiotemporelle pour produire des données de composition et une taille d'échantillon multinomiale qui approxime l'imprécision estimée. Une expérience de simulation qui compare des méthodes spatiotemporelles et basées sur le schéma d'évaluation fait ressortir une augmentation de 32 % de la taille de l'échantillon intrant pour l'estimateur spatiotemporel. Une évaluation avec Stock Synthesis utilisée pour la gestion de la morue-lingue (*Ophiodon elongatus*) dans le courant de Californie montre également une augmentation de 17 % de la taille de l'échantillon et un meilleur calage du modèle en utilisant l'estimateur spatiotemporel, ce qui se traduit par des écarts quadratiques moyens plus faibles pour les estimations de la biomasse de géniteurs. Nous concluons que les approches spatiotemporelles sont viables pour estimer des données d'indice d'abondance et de composition, fournissant ainsi une approche unifiée pour générer des intrants pour les évaluations de stocks. Nous postulons que les méthodes spatiotemporelles amélioreront la puissance statistique pour les données de composition dans de nombreuses évaluations de stocks et recommandons que des travaux futurs se penchent sur l'incidence de l'inclusion d'autres covariables associées à l'habitat ou à l'échantillonnage. [Traduit par la Rédaction]

Introduction

Fish production is highly variable over space and time, and variation is driven by both human and natural causes. Scientific advice regarding fish status and productivity is often based upon population-dynamics models fitted to available data (termed stock assessments). These stock assessments generally use a statistical framework to integrate data that are indexed by time (Maunder and Punt 2013), including a measurement of population abundance (abundance indices), the proportion of population abundance in different age–size–sex categories (composition data), and total fishery removals (catch data) in one or more years.

However, the abundance-index, compositional, and catch data fitted in stock assessments are not directly measured in any single field sample. Instead, fisheries scientists record data obtained from fishers or from preplanned survey programs, and these re-

cords must be aggregated across space to generate inputs that can be fitted by a stock assessment model. For example, data from port-side intercept sampling, fisher logbooks, and observers located on fishing vessels are frequently collected and analysed to estimate total fishery catch (Cotter and Pilling 2007), and these estimates are then treated as catch data in a subsequent stock assessment. Similarly, fishery-independent sampling programs frequently conduct sampling (e.g., bottom trawl tows) at randomized locations, sort the catch to species, weigh biomass for each species, subsample individuals, and then measure length, age, and (or) sex for subsampled individuals (Stauffer 2004). These data are then analysed to estimate abundance-index (using biomass from each sample) and (or) compositional data (using length–age–sex subsampling from each sample), and these estimates are treated as data in an assessment model.

J.T. Thorson* and M.A. Haltuch. Fisheries Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 2725 Montlake Boulevard East, Seattle, WA 98112, USA.
Corresponding author: James Thorson (email: James.Thorson@noaa.gov).

Methods for analysing fishery-dependent and -independent samples to generate abundance-index, compositional, catch, and diet data for use in stock assessments have become more sophisticated over time. For example, hierarchical models have been developed to estimate catch data using fishery logbooks and observer data (Shelton et al. 2012), and model-based analysis of fishery-dependent catch rates decreased bias (Ye and Dennis 2009) and imprecision (Thorson et al. 2017a) relative to simpler analysis methods. In particular, spatiotemporal analysis of survey catch rates resulted in improved precision for simulated data and 28 real-world data sets (Thorson et al. 2015b), and spatiotemporal methods similarly improved precision relative to nonspatial methods when estimating diet proportions (Binion-Rock et al. 2019).

Despite the development of model-based approaches to abundance-index or fishery catch data, there is relatively little prior research regarding model-based estimates of compositional data for use in stock assessments. This is particularly surprising, given the large literature regarding sophisticated statistical approaches when fitting compositional data within an assessment model (Francis 2011; Maunder 2011; Thorson et al. 2017b). Model-based analysis of compositional data could be useful for at least three reasons:

1. Estimate input sample sizes. Analysing compositional data using a statistical model would give an estimate of imprecision as well as the estimated proportion for each length–age–sex, and this imprecision could be used to calculate the "sample size" for the resulting estimate (Thorson 2014). This sample size could then be used as a starting point (or upper bound) during stock assessment data weighting (Francis 2011; Thorson et al. 2017b). Although bootstrapping could instead be used without an explicit statistical model to estimate an input sample size (Stewart and Hamel 2014), a bootstrap analysis typically depends upon the independence of available data and therefore can be biased if this assumption is violated. Although the bootstrap can be extended to correlated data (Fortin and Jacquez 2000), there are many different methods to do so, and an appropriate method has not been proposed for fisheries compositional data.
2. Account for covariates. A statistical model would allow the analyst to explore the potential importance of covariates. In particular, many designs for collecting compositional data include observations at different times and locations, using multiple gears and sampling vessels. Vessel ID is increasingly included as a fixed or random factor when analysing abundance index data (Helser et al. 2004), and season has been used in catch-data analysis (Shelton et al. 2012). Although spatial strata are frequently used when analysing age–length–sex composition records, it is unclear how season, vessel, or other covariates could be included without developing a model-based approach.
3. Improved statistical properties. A model-based framework will often decrease bias, increase precision, and ensure that an estimated confidence interval includes the true value at an intended rate (termed "confidence interval coverage"). For example, accounting for outliers in survey catch rates can improve precision for abundance indices (Thorson et al. 2012), spatiotemporal models can estimate covariance among different sizes or ages arising from the behaviour of a fishing fleet or design of a survey (Berg et al. 2014), and models for multispecies catch rates can be used to control for vessel targeting in fishery catch-rate records (Stephens and MacCall 2004; Winker et al. 2013; Thorson et al. 2017a; Okamura et al. 2017).

These motivations clearly overlap (i.e., accounting for covariates can improve precision), and there could be other reasons to motivate model-based compositional analysis that are not listed.

Given these potential benefits for model-based analysis of compositional records, we develop and demonstrate a spatiotemporal approach to estimate length–age composition data for use in stock assessment. This approach involves (*i*) estimating population density for each length–age–sex category at multiple locations in each year, (*ii*) summing this estimate across space, (*iii*) standardizing by total abundance to estimate proportions, (*iv*) estimating imprecision for proportions, and (*v*) calculating a multinomial sample size that approximates this estimated imprecision. To implement these steps, we modify an existing R package VAST (www.github.com/James-Thorson/VAST) for multivariate spatiotemporal models, which has previously been used for index standardization in stock assessments in the Pacific and North Pacific fisheries management regions. We then use a simulation experiment to compare bias, imprecision, and confidence interval coverage for this spatiotemporal approach with a conventional design-based estimator. Finally, we use bottom trawl survey operations from 2003 to 2015 to generate length-composition data for the stock assessment for lingcod (*Ophiodon elongatus*) in the northern portion of the US California Current to show that the method is feasible to implement in conjunction with existing stock assessment software and increases assessment precision in this case study.

Methods

Expansion of composition data

Age- and size-structured stock assessment models in North America (e.g., Stock Synthesis; Methot and Wetzel 2013) frequently incorporate two forms of "data" arising from these survey operations: an index of abundance $I(t)$ for each year t and a matrix of compositional data $P_c(t)$ for each year t and category c (whether age, size, sex, etc.). In both cases, these data are in fact derived from operations-level samples, and the operations-level data have rarely been included directly within a stock assessment model (although exceptions exist; e.g., Maunder and Langley 2004 or Kristensen et al. 2014). In the following, we focus on abundance-indices and compositional data estimated from fishery-independent survey operations, but note that similar issues arise when processing fishery-dependent records. We also use the term "bins" to refer to categories defined by the size or age of sampled individuals (e.g., length bins for categories defined based on individual length).

Parameters in stock assessment models are then estimated by maximizing their log-likelihood given available data. Specifically, a probability density is specified for abundance-index data given assessment estimates of available biomass:

$$(1) \qquad I(t) \sim \text{Lognormal}\left(\sum_{c=1}^{n_c} q S_c W_c N_c(t), \, \sigma^2(t) \right)$$

where q is an estimated catchability coefficient, S_c is the estimated selectivity for each bin, W_c is the biomass-per-individual for bin c (which may be estimated or input as data), $N_c(t)$ is estimated abundance for each bin and year, and $\sigma^2(t)$ is the variance for the abundance index in year t, which is input as data (see Table 1 for a list of notation). Similarly, a probability distribution is specified for compositional data $P(t)$ representing the proportion of abundance (or biomass) $P_c(t)$ within category c, given assessment estimates $\pi(t)$ of the proportion of available population abundance (or biomass) $\pi_c(t)$ for each age–length–sex bin c:

$$(2) \qquad P(t) \sim \text{Multinomial}(\pi(t), \tau(t))$$

Table 1. Names and symbols for variables and data as used in the main text, as well as whether the variable is used in the theoretical explanation ("Exp.", eqs. 1–3), estimation model (EM, eqs. 4–13), the operating model used to simulate data for model testing (OM, eqs. 14–17), or model diagnostics ("Diag.", eqs. 18–19).

Name	Symbol	Use
Abundance index	$I(t)$	Exp.
Catchability coefficient in assessment model	q	Exp.
Selectivity at age in assessment model	S_c	Exp.
Mass-at-age in assessment model	W_c	Exp.
Numbers-at-age in assessment model	N_c	Exp.
Abundance index variance	$\sigma^2(t)$	Exp.
Proportion-at-age in assessment model	$\pi_c(t)$	Exp.
Observed proportion-at-age	$P_c(t)$	Exp.
Observed biomass in sample i	b_i	Exp.
Observed number of individuals in sample i	n_i	Exp.
Subsampled biomass in sample i	$\tilde{b}_i$	Exp.
Subsampled abundance-at-age in sample i	$\tilde{n}_c(i)$	Exp.
Subsampling intensity for sample i	λ_i	Exp.
Abundance-at-age for sample i after first-stage expansion	$n_c(i)$	Exp., EM
Predicted encounter probability	$p_c(i)$	OM, EM
Expected biomass when encountered	$r_c(i)$	OM, EM
Area swept for sample i	a_i	EM
Area associated with knot s	$a(s)$	EM
Predicted density for category, location, and time	$d_c(s, t)$	EM
Dispersion for probability density function for positive catch rates	$\sigma_M(c)$	EM
Intercepts	$\beta_p(c, t)$	EM
Spatial variation for encounter probability r	$\omega_p(c, s)$	EM
Spatiotemporal variation for encounter probability r	$\varepsilon_p(c, s, t)$	EM
Standard deviation for spatial variation	$\sigma_{\omega p}(c)$	EM
Standard deviation for spatiotemporal variation	$\sigma_{\varepsilon p}(c)$	EM
Predicted proportion for each category and year	$\hat{P}_c(t)$	EM
Predicted index of population abundance for category c	$\hat{I}_c(t)$	EM
Predicted index of population biomass across all categories	$\hat{I}(t)$	EM
Input sample size for predicted proportions	$\hat{\tau}(t)$	EM
Average recruitment	β_N	OM
Spatial variation in recruitment	$\omega_N(s)$	OM
Spatiotemporal variation in recruitment	$\varepsilon_N(s)$	OM
Annual instantaneous survival rate	Z	OM
Abundance at age for each cell	$N_a(s, t)$	OM
Expected mass-at-age	$W_a(s, t)$	OM
Brody growth coefficient	K	OM
Asymptotic maximum length	L_∞	OM
Tissue density	w_α	OM
Allometric scaling of mass and length	w_β	OM
Spatial variation in mass-at-age	$\omega_W(s)$	OM
Spatiotemporal variation in mass-at-age	$\varepsilon_W(s, t)$	OM
Error in estimated proportion-at-age	$E_c(t)$	Diag.
Likelihood-ratio statistic for estimated proportion-at-age and sample size	$\chi^2(t)$	Diag.
Quantile for estimated proportion-at-age and sample size	$Q(t)$	Diag.

where $\pi_c(t) = \dfrac{S_c N_c(t)}{\sum_{c=1}^{n_c} S_c N_c(t)}$ is the proportion of abundance in each bin c in year t that is estimated by the stock assessment model, and $\tau(t)$ is the sample size for compositional data in year t (which is input as data). Both equations are frequently modified to incorporate "overdispersion" (e.g., instances where the fit between abundance index or compositional data has greater variance than would be expected given the abundance variance $\sigma^2(t)$ or compositional input sample size $\tau(t)$). In these cases, the assessment model frequently estimates additional overdispersion, and this is equivalent to estimating the relative weight associated with each data source in the assessment model (Francis 2011; Thorson et al. 2017b).

Surveys for sampling marine populations are frequently designed to collect samples in each year at locations that are chosen using simple or stratified random sampling. At each chosen location, a gear is deployed and results are recorded. In the following, we refer to bottom trawl sampling, but results could also apply to other samples at discrete locations (e.g., longline or hook-and-line sampling gears). Gear operations typically yield multiple records for each tow i: (i) the total biomass b_i and (or) abundance n_i for a given species; (ii) the biomass $\tilde{b}_i$ that is subsampled for a given species; and (iii) attributes (e.g., length, mass, age, sex, etc) measured from individuals that are subsampled from the set of captured individuals for each species. Individual-specific attributes are then aggregated to calculate the abundance $\tilde{n}_c(i)$ for different age–length–sex categories in the subsample, and records are analyzed to generate inputs to stock assessment models ($I(t)$ and $P_c(t)$).

The proportion of total biomass that is subsampled for ages or lengths, termed the "subsampling intensity" λ_i (where $\lambda_i = \tilde{b}_i/b_i$ or $\lambda_i = \tilde{n}_i/n_i$ depending upon whether b_i or n_i is measured) may vary between tows. If subsampling intensity is correlated with total biomass b_i, then the subsampling design is not "ignorable" and must be included during analysis to yield unbiased estimates of compositional data. For this reason, subsamples are typically expanded to represent total abundance in each survey tow, and we refer to this step as "first-stage expansion" (Thorson 2014). This first-stage expansion involves predicting abundance $n_c(i)$ for each

404

Can. J. Fish. Aquat. Sci. Vol. 76, 2019

bin as the ratio of subsampled proportions and the subsampling intensity $n_c(i) = \tilde{n}_c(i)/\lambda_i$ and then analysing the expanded abundance $n_c(i)$ using a statistical model.

Design-based approach to expanding compositional data

Predicted abundance per bin $n_c(i)$ after first-stage expansions is then typically expanded to the total area sampled by a given survey design, and we call this "second-stage expansion." The second-stage expansion is typically conducted using a "design-based" algorithm, where the predicted abundance $n_c(i)$ is summed across all records in a given stratum to calculate the total number per stratum. Total number per stratum is then averaged across strata with each stratum weighted by its total area (for survey data) or landings (for fishery data), then these aggregate numbers are normalized to calculate a proportion $P_t(c)$ for each bin c and year t. Given simple-random sampling (i.e., only one sampling stratum), this second-stage expansion reduces to

$$(3) \qquad \hat{P}_c(t) = \frac{\sum_{i=1}^{n_i} n_c(i)}{\sum_{c=1}^{n_c} \sum_{i=1}^{n_i} n_c(i)}$$

where $\hat{P}_c(t)$ is then fitted as data in the subsequent stock assessment model.

The spatiotemporal approach to analysing compositional data

We propose an alternative approach to estimating proportions $P_t(c)$ using a vector autoregressive spatiotemporal (VAST) model (Thorson and Barnett 2017c). This alternative builds upon Berg et al. (2014), which applied a generalized additive model (GAM) with a single spatial smoother and annually varying intercept to estimate an abundance index $I_c(t)$ for each age, where this age-specific abundance index was then fitted directly to abundance-at-age $N_c(t)$ within an assessment model. Fitting abundance-at-age and its standard errors $SE[N_c(t)]$ within an assessment model is common in European assessment models (e.g., Nielsen and Berg 2014; Albertsen et al. 2016). By contrast, we estimate proportions $P_t(c)$, input sample size $\tau(t)$, and total abundance $I(t)$ as is common for assessment models in the US West Coast (e.g., Methot and Wetzel 2013).

We specifically fit a spatiotemporal delta model to expanded numbers per bin $n_c(i)$

$$(4) \qquad \Pr(n_c(i) = B) = \begin{cases} 1 - p_c(i) & \text{if } B = 0 \\ p_c(i) \times \text{Lognormal}\left(B \mid r_c(i), \sigma_m^2(c)\right) & \text{if } B > 0 \end{cases}$$

where $p_c(i)$ is the predicted encounter probability, $r_c(i)$ is the predicted log-biomass if encountered, $\text{Lognormal}(B \mid r_c(i), \sigma_m^2(c))$ is a lognormal probability density function, and $\sigma_m^2(c)$ governs the predicted variance for positive catch rates. This model involves linear predictors for $p_c(i)$ and $r_c(i)$:

$$(5) \qquad \begin{aligned} \text{logit}[p_c(i)] &= \beta_p(c, t_i) + \sigma_{\omega p}(c)\omega_p(c, s_i) + \sigma_{\varepsilon p}(c)\varepsilon_p(c, s_i, t_i) \\ \log[r_c(i)] &= \log(a_i) + \beta_r(c, t_i) + \sigma_{\omega r}(c)\omega_r(c, s_i) + \sigma_{\varepsilon p}(c)\varepsilon_r(c, s_i, t_i) \end{aligned}$$

where $\beta_p(c, t_i)$ and $\beta_r(c, t_i)$ are intercepts for each bin c and year t, $\omega_p(c, s_i)$ and $\omega_r(c, s_i)$ are random effects representing spatial variation among locations s_i for each sample i (e.g., a tow, visual sample, etc.), $\varepsilon_p(c, s_i, t_i)$ and $\varepsilon_r(c, s_i, t_i)$ are random effects representing spatiotemporal variation among locations s_i and times t_i for each sample, and $\log(a_i)$ is the log-area sampled as a linear offset for $\log[r_c(i)]$. We specify a distribution for spatial and spatiotemporal random effects:

$$(6) \qquad \begin{aligned} \omega_p(c) &\sim \text{MVN}(\mathbf{0}, \mathbf{R}_p) \\ \omega_r(c) &\sim \text{MVN}(\mathbf{0}, \mathbf{R}_r) \\ \varepsilon_p(c, t) &\sim \text{MVN}(\mathbf{0}, \mathbf{R}_p) \\ \varepsilon_r(c, t) &\sim \text{MVN}(\mathbf{0}, \mathbf{R}_r) \end{aligned}$$

where $\mathbf{R}_p$ and $\mathbf{R}_r$ are correlation matrices for $p_c(i)$ and $r_c(i)$, respectively, as calculated from a Matern function:

$$(7) \qquad \mathbf{R}_p(s, s + h) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \times (\kappa_p|h\mathbf{H}|)^{\nu} \times K_{\nu}(\kappa_p|h\mathbf{H}|)$$

where $\mathbf{H}$ is a linear transformation representing geometric anisotropy that is assumed to be identical for $\mathbf{R}_p$ and $\mathbf{R}_r$ and that involves estimating two parameters (the direction of the major axis of geometric anisotropy and the ratio of major and minor axes), ν is the smoothness of the Matern function (fixed at 1.0), and κ_p governs the decorrelation distance for $\mathbf{R}_p$ and where κ_r is separately estimated for $\mathbf{R}_r$ (see Thorson et al. 2015b for more details). We specify that random effects $\omega_p(c)$, $\omega_r(c)$, $\varepsilon_p(c, t)$, and $\varepsilon_r(c, t)$ have a standard deviation of 1.0, so that parameters $\sigma_{\omega p}(c)$, $\sigma_{\varepsilon p}(c)$, $\sigma_{\omega r}(c)$, and $\sigma_{\varepsilon p}(c)$ are identifiable and are interpreted as the standard deviation of a given spatial or spatiotemporal process.

Fixed effects ($\beta_p(c, t)$, $\sigma_{\omega p}$, $\sigma_{\varepsilon p}$, $\beta_r(c, t)$, $\sigma_{\omega r}$, $\sigma_{\varepsilon r}$, and $\sigma_m(c)$) are estimated by identifying their values that maximize the marginal log-likelihood function. The marginal log-likelihood is calculated by integrating the joint log-likelihood of data and random effects with respect to random effects ($\omega_p(c, s)$, $\varepsilon_p(c, s, t)$, $\omega_r(c, s)$, and $\varepsilon_r(c, s, t)$). We use Template Model Builder (TMB; Kristensen et al. 2016) to implement the Laplace approximation to the marginal log-likelihood (Skaug and Fournier 2006), and TMB uses automatic differentiation to also calculate the gradients of the marginal log-likelihood with respect to fixed effects. We approximate the probability of spatial random effects using the SPDE method (Lindgren et al. 2011). Preliminary exploration suggests that age–length–sex categories with few positive encounters can result in poor convergence for bin-specific hyperparameters (i.e., $\sigma_{\omega p}(c)$, $\sigma_{\varepsilon p}(c)$, $\sigma_{\omega r}(c)$, $\sigma_{\varepsilon r}(c)$, and $\sigma_m^2(c)$). We therefore impose the restriction that these parameters are equal for any bin with fewer than X encounters combined throughout all observed years (where the value for X is chosen based on model exploration). For identifiability, we also "turn off" all intercepts for any combination of year t and category c with no encounters (i.e., $n_c(i) = 0$ for all i) and fix $\beta_p(c, t)$ to a sufficiently low value such that predicted encounter probabilities p_c approach zero for that year.

We identify the maximum likelihood estimate (MLE) of fixed effects using a gradient-based nonlinear minimizer within the R statistical environment (R Core Team 2016). Specifically, we use a gradient-based version of the Nelder–Mead algorithm five times, each time starting from the MLE identified the previous iteration. At the end of each Nelder–Mead iteration, we identify any spatiotemporal hyperparameter that approaches zero (i.e., <0.001) and fix these parameters to zero. We eliminate hyperparameters that approach zero because these parameters are approaching a boundary in parameter space, and having parameters at boundaries will generally complicate any interpretation of asymptotic standard errors. We also use a final Newton-step (based on the Hessian matrix after the final Nelder–Mead iteration) to tighten convergence and confirm that the absolute gradient of the marginal log-likelihood with respect to each fixed effect is <0.001.

Estimated values of fixed and random effects are then used to predict density $\hat{d}_c(s,t)$ for every location s and year t:

$$(8) \qquad \hat{d}_c(s, t) = \hat{p}_c(s, t) \times \hat{r}_c(s, t)$$

where density $\hat{d}_c(s,t)$ is then used to predict total abundance for the entire domain (or a subset of the domain) for a given species:

$$(9) \qquad \hat{I}_c(t) = \sum_{s=1}^{n_s} [a(s) \times \hat{d}_c(s,t)]$$

where $a(s)$ is the area associated with location s (Thorson et al. 2015b). We then use this index to calculate total abundance:

$$(10) \qquad \hat{I}(t) = \sum_{c=1}^{n_c} \hat{I}_c(t)$$

the proportion for each bin:

$$(11) \qquad \hat{P}_c(t) = \frac{\hat{I}_c(t)}{\hat{I}(t)}$$

and the estimation variance $\widehat{SE}[\hat{P}_c(t)]^2$ for proportions:

$$(12) \qquad \widehat{SE}[\hat{P}_c(t)]^2 \approx \frac{\hat{I}_c(t)^2}{\hat{I}(t)^2} \left\{ \frac{\widehat{SE}[\hat{I}_c(t)]^2}{\hat{I}_c(t)^2} - 2\frac{\widehat{SE}[\hat{I}_c(t)]^2}{\hat{I}(t)\hat{I}_c(t)} + \frac{\sum_{c=1}^{n_c} \widehat{SE}[\hat{I}_c(t)]^2}{\hat{I}(t)^2} \right\}$$

where we use this approximation (see online Supplementary material A for derivation[1]) because it relies only on standard errors for each individual category $(\widehat{SE}[\hat{I}_c(t)]^2)$ and does not require a standard error for any value calculated from multiple categories (e.g., as required to calculate $\widehat{SE}[\hat{I}(t)]^2$). Calculating the standard error for a value derived from multiple categories eliminates the sparsity of Hessian matrix used in the Laplace approximation, so our approximation results in improved computational efficiency. For any category c and year t with no encounters, we specify that $\hat{I}_c(t) = \widehat{SE}[\hat{I}(t)]^2 = 0$, such that model behaviour matches a design-based estimator in these instances. We then use estimated proportions and standard errors to calculate the input sample size $\hat{\tau}(t)$:

$$(13) \qquad \hat{\tau}(t) = \text{Median}_c \left\{ \frac{\hat{P}_c(t)[1 - \hat{P}_c(t)]}{\widehat{SE}[\hat{P}_c(t)]^2} \right\}$$

where this formula is derived by calculating the multinomial sample size that would approximate the estimated imprecision for proportion $\hat{P}_c(t)$ (see Thorson 2014) and where we use the median instead of the mean because that was the version previously published by Thorson (2014). For comparison, we also calculate the input sample size $\hat{\tau}(t)$ for the design-based algorithm, using the sample variance of expanded numbers $n_c(i)$ across samples i in year t divided by the number of samples in that year as $\widehat{SE}[\hat{I}_c(t)]^2$ in eq. 12. Future research could explore passing the estimated covariance $\widehat{Cov}[\hat{P}_c(t)]$ for proportions directly to the assessment model, although we do not explore the idea further here.

After identifying the MLE, TMB uses the generalized delta-method to calculate all standard errors as a function of fixed and random effects (Kass and Steffey 1989). TMB also implements the epsilon-estimator to bias-correct predicted abundance $\hat{I}_c(t)$ to account for the nonlinear transformation of random effects (Thorson and Kristensen 2016), and this bias-corrected prediction of $\hat{I}_c(t)$ is used when calculating $\hat{P}_c(t)$ (eq. 11) and $\hat{\tau}(t)$ (eq. 13). The spatiotemporal model is implemented using version 1.6.0 of R package VAST

(Thorson and Barnett 2017), publicly available at https://github.com/James-Thorson/VAST.

Simulation experiment

We explore the performance of the spatiotemporal model for analyzing compositional data using a simulation experiment with 100 replicates. To do so, we simulate data arising from simple age-structured dynamics and compare estimates of proportion-at-age $\hat{P}_c(t)$ from design and spatiotemporal estimators with its true value $P_c(t)$. Specifically, we simulate abundance (in numbers) at age $N_a(s, t)$ for a short-lived groundfish at each of 58 049 (2 km by 2 km) cells throughout the Eastern Bering Sea:

$$(14) \qquad N_a(s,t) = \begin{cases} \exp[\beta_N + \omega_N(s) + \varepsilon_N(s,t)] \times \exp(-aZ) & \text{if } t = 1 \text{ or } a = 1 \\ N_{a-1}(s, t-1) \times \exp(-Z) & \text{if } t > 1, \ a > 1 \end{cases}$$

where $\exp(\beta_N)$ is the median density for age-0 individuals, $\omega_N(s)$ and $\varepsilon_N(s, t)$ are spatial and spatiotemporal variation in log-density, and $Z = 0.5 \cdot \text{year}^{-1}$ is the instantaneous mortality rate (from natural and human causes). We also simulate variation in mass-at-age $W_a(s, t)$:

$$(15) \qquad W_a(s,t) = w_\alpha \{L_\infty[1 - \exp(-Ka)]\}^{w_\beta} \times \exp[\omega_W(s) + \varepsilon_W(s,t)]$$

where $L_\infty = 100$ cm is asymptotic length, $K = 0.3 \cdot \text{year}^{-1}$ is the Brody growth coefficient, $w_\alpha = 0.01$ g·cm^{-3} is tissue density, $w_\beta = 3$ is isometric scaling of mass-at-length, and $\omega_W(s)$ and $\varepsilon_W(s, t)$ are spatial and spatiotemporal variation, respectively, in mass-at-age. Spatial (ω_N and ω_W) and spatiotemporal ($\varepsilon_N(t)$ and $\varepsilon_W(t)$) terms are simulated from a Gaussian random field (eq. 6) using R package RandomFields (Schlather 2009). However, we use an isotropic Gaussian covariance function in the simulation model in place of the anisotropic Matérn covariance function (eq. 7) used during parameter estimation. We specify in the simulation model a scale of 250 km for all simulated spatial fields (i.e., locations 379 km apart have a 10% correlation), where spatial terms (ω_N and ω_W) have respective standard deviations of 0.5 and 0.1, spatiotemporal terms ($\varepsilon_N(t)$ and $\varepsilon_W(t)$) have respective standard deviations of 0.2 and 0.1, and measurement errors have comparable magnitude to total spatial and spatiotemporal variation ($\sigma_m^2(c) = 1$).

We then simulate sampling at each sampled location and year from the historical Eastern Bering Sea bottom trawl survey for the last 25 years (1992–2016), following a Poisson-link delta-model (Thorson 2018):

$$(16) \qquad \begin{aligned} p_i(a) &= 1 - \exp[-a_i S_a N_a(s_i, t_i)] \\ r_i(a) &= \frac{a_i S_a N_a(s_i, t_i)}{p_i(a)} \times W_a(s_i, t_i) \end{aligned}$$

where a_i is the area swept by sample i occurring at location s_i and year t_i, $S_a = \{1 - \exp[-S_{\text{slope}}(a - S_{0.5})]\}^{-1}$ is selectivity-at-age ($S_{0.5} = 3$ years is age at 50% selection and $S_{\text{slope}} = 1$ governs the slope of the selectivity-at-age curve) and where encounter probability $p_i(a)$ and positive catch rate $r_i(a)$ is defined such that encounter probability follows a complementary log-log distribution and $D_a(s_i, t_i) = p_a(s_i, t_i) r_a(s_i, t_i) = N_a(s_i, t_i) W_a(s_i, t_i)$. We then simulate data using the conventional delta-model (eq. 4) and also calculate the true proportion-at-age:

$$(17) \qquad P_c(t) = \frac{\sum_{s=1}^{n_s} [a(s) \times D_a(s,t)]}{\sum_{a=1}^{n_a} \sum_{s=1}^{n_s} [a(s) \times D_a(s,t)]}$$

where we record this value for each year in each replicate.

---

[1]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2018-0015.

We simulate 100 replicated data sets using this simulation model and fit the spatiotemporal compositional expansion model to each. We then extract predictions of $\hat{P}_c(t)$ and $\hat{\tau}(t)$ for each replicate and evaluate model performance in three ways:

1. Bias and imprecision. We record the true $P_c(t)$ and estimated $\hat{P}_c(t)$ proportion using both spatiotemporal and design-based estimators, calculate the error for each age and year $E_c(t) = \hat{P}_c(t) - P_c(t)$, and calculate the average $E_c(t)$ across years, ages, and replicates (a measure of bias) and standard deviation of $E_c(t)$ (a measure of imprecision);

2. Confidence interval coverage. We record the predicted input sample size $\hat{\tau}(t)$ for both spatiotemporal and design-based estimators and use a likelihood-ratio test for the null hypothesis that $\hat{P}_c(t)\hat{\tau}(t)$ in each year t arises from a multinomial distribution with proportion $P_c(t)$ and sample size $\hat{\tau}(t)$. This test involves calculating a likelihood-ratio statistic χ_t^2 for every year of each simulation replicate:

$$(18) \qquad \chi^2(t) = \sum_{c=1}^{n_c} \hat{\tau}(t)\hat{P}_c(t) \log\left[\frac{\hat{P}_c(t)}{P_a(t)}\right]$$

We then compare this test statistic with its theoretical distribution under the null hypothesis, which follows a χ^2 distribution with n_c degrees of freedom. Given this theoretical distribution, we extract the quantile $Q(t)$ of the test statistic $\chi^2(t)$ from a χ^2 distribution with n_a degrees of freedom and record this quantile for every year and replicate for both estimators. $Q(t)$ should have a uniform distribution under the null hypothesis that estimated input sample size $\hat{\tau}_t$ is a good summary of the variance between predicted $\hat{P}_c(t)$ and true $P_c(t)$ proportions, while excess mass as $Q(t) \to 0$ implies that the input sample size $\hat{\tau}(t)$ is too small (i.e., $\widehat{SE}[\hat{P}_c(t)]^2$ is too large), while excess mass as $Q(t) \to 1$ implies that $\hat{\tau}(t)$ is too large.

3. Estimated sample size. Finally, we record the estimated sample size $\hat{\tau}(t)$ for each year and replicate for both design ($\hat{\tau}_{design}(t)$) and model-based estimators ($\hat{\tau}_{model}(t)$). We then calculate the average ratio under the assumption that estimated sample size follows a skewed (i.e., lognormal) distribution:

$$(19) \qquad \log[\hat{\tau}_{model}(t)] \sim \text{Normal}\left(\mu + \log[\hat{\tau}_{design}(t)], \sigma_\tau^2\right)$$

and we report the value $\exp(\mu) - 1$ (and its standard error) as a summary of the percent change in sample size from using a spatiotemporal estimator relative to the conventional design-based estimator.

We expect that both design- and model-based estimators will be approximately unbiased and have a similar distribution for quantiles $Q(t)$. Based on previous comparisons of index-standardization methods (e.g., Thorson et al. 2015b), we also expect that the model-based estimator should have smaller imprecision and larger estimated sample sizes than the design-based estimator. This expected increase in precision would occur if the model-based estimator explains some portion of observed variance via spatiotemporal patterns and therefore has a smaller estimate of residual variance when predicting proportions into unsampled locations.

Case study

We also demonstrate that the spatiotemporal expansion of compositional data are feasible using real-world data for the stock of lingcod (*Ophiodon elongatus*) off the Oregon and Washington coasts. We specifically compare stock assessment model performance when using a design-based or a spatiotemporal model-based expansion of length-composition data in the 2017 stock assessment model (Haltuch et al. 2017). To do so, we conduct both

design-based and model-based expansion of length-composition records from the Northwest Fisheries Science Center (NWFSC) shelf-slope bottom trawl survey (Keller et al. 2017), the primary source of fishery-independent information for US West Coast groundfishes. Following the approach used in the 2017 assessment, we expand compositional estimates to the Oregon and Washington portions of the sampling domain (from 42°N to 49°N), which corresponds to 12 017 (2 km by 2 km) cells.

For simplicity of presentation, we make the following changes to the stock assessment model:

1. We separately expand and fit to data for male and female individuals (rather than simultaneously expanding them). This change is intended to simplify visualization of results when comparing design and model-based estimates of proportion-at-length.

2. We include newly expanded length-composition data for the US West Coast bottom trawl survey from 2013 to 2015. We also include input sample sizes calculated for design-based ($\hat{\tau}_{design}(t)$) and model-based ($\hat{\tau}_{VAST}(t)$) estimators.

3. We use the Dirichlet-multinomial distribution to re-estimate the effective sample size given the input sample sizes for each method (Thorson et al. 2017b), where this sample size represents the amount of information in expanded compositional data as estimated by the stock assessment model (i.e., how well compositional data match other data sources and model assumptions).

We specifically estimate $\hat{P}_c(t)$ for 13 years and 60 length bins (from 10 to 130 cm length divided into 2 cm bins) and use 100 spatial knots to approximate spatial resolution (i.e., using the same spatial resolution as used by Thorson et al. (2015b) for abundance index testing). The stock assessment is implemented in Stock Synthesis (Methot and Wetzel 2013) and is otherwise identical to that recently accepted for management off the US West Coast (see online Supplementary material B for a summary of the model[1]).

We implement first-stage expansion using the ratio of sampled and total numbers (i.e., $\lambda_i = \tilde{n}_i/n_i$) and then analyse resulting records $n_c(i) = \tilde{n}_c(i)/\lambda_i$ using a conventional delta-model. Some length bins have too few observations to reliably estimate hyperparameters $\sigma_m(c)$, $\sigma_{\varepsilon p}(c)$, and $\sigma_{\varepsilon p}(c)$. We therefore estimate a single value for each hyperparameter for all length bins that have fewer than 100 observations in total from 2003 to 2015. We then record estimates of length-composition data $\hat{P}_c(t)$ for both design- and model-based estimators and include both in the 2017 stock assessment model. We record estimates of the biomass ratio (spawning output relative to its average unfished level) and fishing exploitation rate (spawning potential ratio) for each assessment model run and compare these values as well as the value of the negative log-likelihood using each data set.

Results

Comparison between spatiotemporal and design-based estimates of proportion-at-age for a replicate of the simulation experiment (Fig. 1) shows broad agreement between methods, where both are able to distinguish years with relatively strong old (year 14) or young (year 12) cohorts, as well as track cohorts (e.g., the cohort born in year 4 is dominant from ages 3 to 7 in years 7 to 11). The estimated sample size $\hat{\tau}(t)$ in this replicate is generally larger for the spatiotemporal than the design-based estimator (e.g., 115 versus 68 in year 1). Comparing the error between spatiotemporal and design-based estimators (Fig. 2) confirms that both methods are approximately unbiased (both have an average bias within <0.2%). However, the spatiotemporal estimator has average error that is approximately 10% smaller than the design-based estimator in every age.

Both spatiotemporal and design-based estimators overestimate sample size $\hat{\tau}(t)$ by a similar amount. In particular, the quantile

Fig. 1. Prediction of proportion-at-age $\hat{P}_a(t)$ (y axis) for each age (x axis) in each simulated year (panel) using a design-based (blue) or spatiotemporal model-based (red) estimator (solid line: predicted value; vertical whisker: ±1 standard error), including the input sample size for design-based $\tau_{design}(t)$ (blue number) and spatiotemporal models $\tau_{VAST}(t)$ (red number), compared with true proportion-at-age $P_a(t)$ (black) for a single replicate of the simulation experiment. [Colour online.]
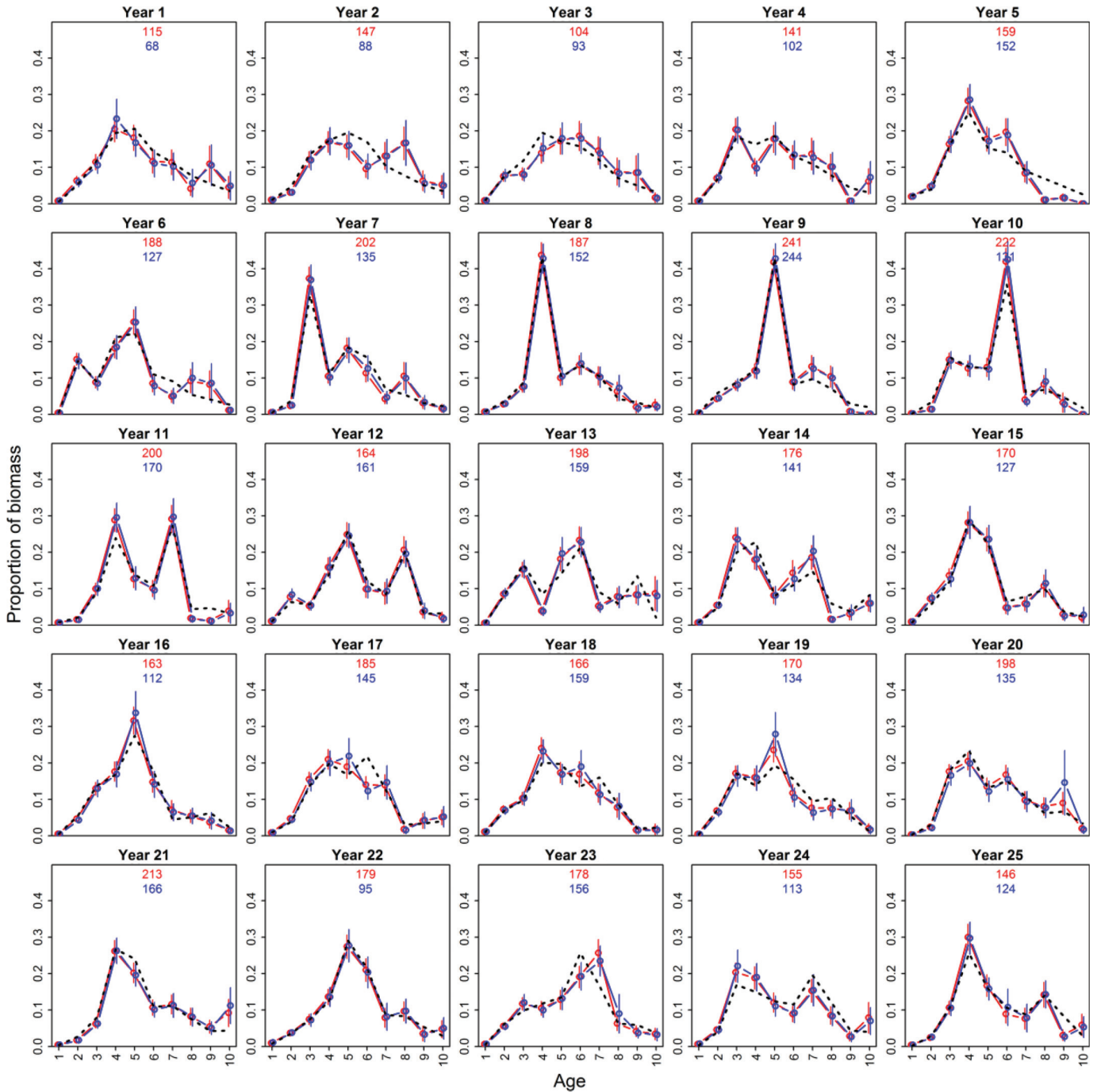
$Q(t)$ for both models (Fig. 3) indicates that the true proportion-at-age $P_a(t)$ is outside the 90% predictive interval approximately 40% of the time (i.e., the quantile $0.9 < Q(t) < 1.0$ contains almost 35%–40% of replicates for both models). This indicates that the estimated sample size $\hat{\tau}(t)$ is greater than it should be for about 25%–30% of replicates and is appropriate for the remaining 70%–75%. We therefore conclude that sample size estimates are satisfactory in nearly 75% of simulation replicates and are overestimated in the remaining 25% and that performance is similar between design and model-based estimators. Comparison of estimated sample

size $\hat{\tau}(t)$ between estimators (Fig. 4) indicates that the spatiotemporal model yields proportion-at-age estimates $\hat{P}_a(t)$ with a 32% greater sample size than the design-based estimator. However, individual years and replicates result in a lower spatiotemporal sample size than the design-based estimator (i.e., dots below the dashed one-to-one line in Fig. 4).

Finally, applying both spatiotemporal and design-based estimators to length-structured data for lingcod in the US California Current (Fig. 5) shows that both methods generate very similar estimates of proportion-at-size. For example, a strong cohort first
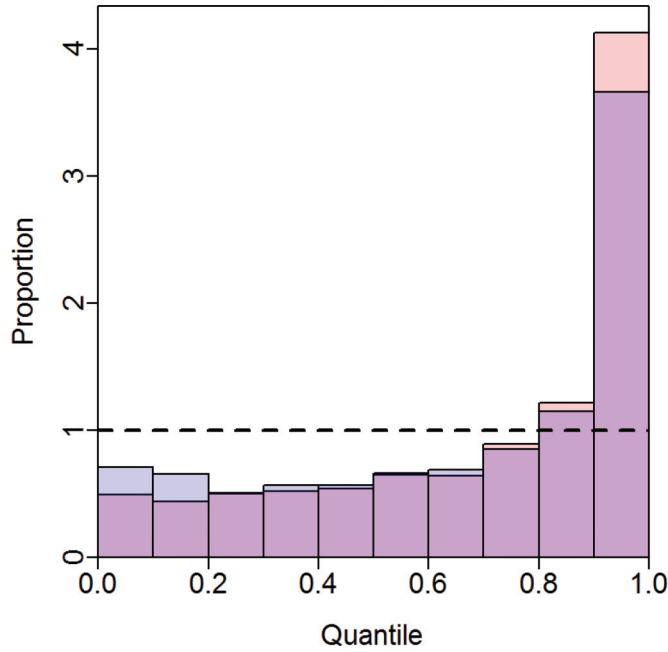
408

Can. J. Fish. Aquat. Sci. Vol. 76, 2019

Fig. 2. Summary of error ($[\hat{P}_c(t) - P_c(t)] \times 100\%$) in the simulation experiment (y axis) for each year (x axis) and age (panel), showing average (solid line) and ±1 standard deviation for error (shaded interval) for the spatiotemporal (red) and design-based estimator (blue), where each panel includes the average error across years with the average root-mean-squared error across years in parenthesis. A well-performing model will have zero average error and will minimize the root-mean-squared error. [Colour online.]
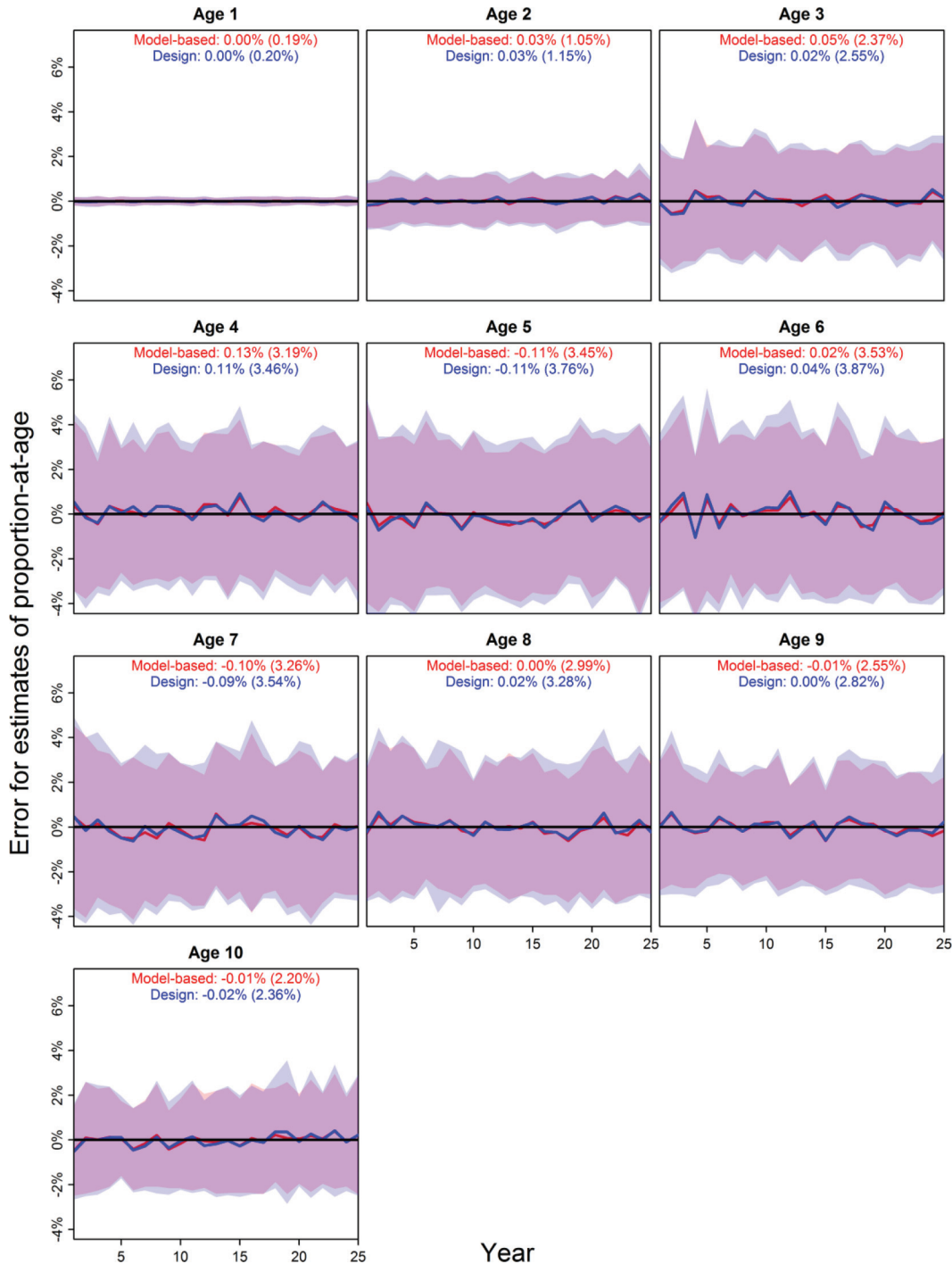
appears in 2009 (at a size of approximately 26 cm), and this strong cohort can be easily identified until 2014 in both male and female composition data. However, small differences do arise; for example, the increase in 42–44 cm females in 2003 in the model-based relative to design-based estimates and vice-versa in 54–56 cm females in 2008. Model diagnostics (Supplementary Fig. C1[1]) show that encounter probability is shrunk towards its average value for both males and females and therefore is underestimated for the small proportion of observations with high encounter propor-

tion, but otherwise do not indicate a lack of model fit. In general, the model-based approach yields larger input-sample sizes ($\hat{\tau}(t)$ from eq. 13, averaged across years and sexes: 211 for design-based and 248 for model-based estimators). When fitting to these data sets and re-estimating effective sample size, the model-based approach results in a larger effective sample size, indicating a small increase in goodness of fit for expanded compositional data when using the model-based approach (Fig. 6, top panel). The larger effective sample size in turn underlies a small decreased in confi-

Fig. 3. Proportion of simulated replicates and years (*y* axis) with predictive quantiles $Q(t)$ within a given range (*x* axis) for the spatiotemporal model (red) or design-based (blue) estimators (purple shows overlap between model and design-based estimators), where a perfect model would have a quantile distribution of 1.0 (dashed horizontal line). [Colour online.]
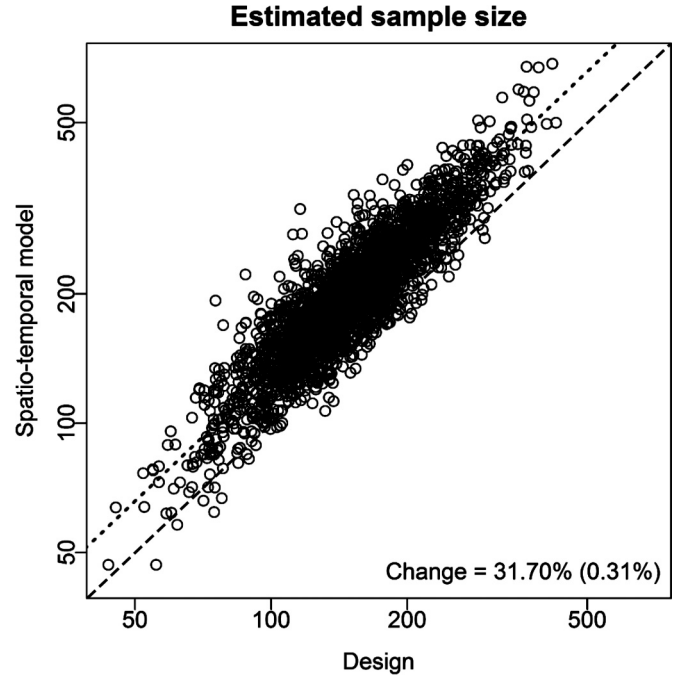

Fig. 4. Input sample size for design-based $\tau_{design}(t)$ (*x* axis) and spatiotemporal models $\tau_{VAST}(t)$ (*y* axis) for each simulation replicate and year (circles), showing the 1:1 line (dashed line) representing equal sample size and the estimated ratio of model versus design-based sample size (dotted line) and the average ratio (bottom right) and its standard error expressed as a percentage (see eq. 19).

Estimated sample size

Change = 31.70% (0.31%)

dence interval width for spawning biomass when fitting to the model-based compositional data, but no substantial change in the maximum likelihood estimate for spawning biomass or status relative to unfished levels.

Discussion

We used a simulation experiment and real-world case study to demonstrate a spatiotemporal approach to estimating compositional data for use in stock assessment models. The simulation experiment showed that a spatiotemporal model substantially improves precision while remaining unbiased and having satisfactory confidence interval coverage. An application to lingcod in the US California Current confirmed that our implementation is feasible using real-world data in conjunction with the existing Stock Synthesis assessment software, even given the small (2 cm) length bins that are commonly widely used on the US West Coast (Monnahan et al. 2016). The case study also corroborates the conclusion that a spatiotemporal approach can increase both input and effective sample size for compositional data. Our implementation involves using R package VAST to estimate spatiotemporal variation in density for each age–length–sex bin, and this package has previously been used for stock assessment in the Pacific and North Pacific fisheries management councils (e.g., Lunsford et al. 2015; Thorson and Wetzel 2015). We therefore believe that our proposed approach could be feasible for use in real-world assessments.

We envision that estimating input sample size has an important role in estimating the relative information in composition and abundance-index data. Determining the relative weighting to give these two data types during stock assessment model development has been called "data weighting" (Francis 2011) and remains a hotly contested topic in stock assessment research (see e.g., Maunder et al. 2017). Our spatiotemporal model for estimating composition data calculates the multinomial sample size $\hat{\tau}(t)$ that approximates the estimated precision for proportions $\hat{P}_c(t)$

(i.e., adapting methods from Thorson 2014). This multinomial sample size could then be treated as the input sample size when fitting composition data in an assessment model, as we do in the case study here. However, the variance between expected and observed proportions will often exceed its expectation given the input sample size, which we call "overdispersion". Overdispersion can arise due to model mis-specification (e.g., because the assessment model neglects spatial processes causing changes in selectivity), such that the input sample size is too high. We therefore recommend estimating overdispersion of compositional data within every assessment, using the multinomial sample size as an upper bound on effective sample size. This approach is computationally efficient using the Dirichlet-multinomial distribution (Thorson et al. 2017b), although the Dirichlet-multinomial has the same correlation structure as the multinomial distribution and may perform poorly when overdispersion is correlated among ages (e.g., due to mis-specified selectivity assumptions). If this estimated overdispersion is large, we foresee that the analyst could increase model flexibility for that model component, for example, by allowing selectivity to vary over time (Xu et al., 2019) as it is often observed to do (Sampson and Scott 2012). We therefore interpret model-based estimates of input sample size as an important starting point in an iterative approach to stock assessment model development and call this a "model-based approach to compositional data weighting" (see Fig. 7 for a visualization).

Ongoing research has developed assessment models that fit directly to fishery-dependent and -independent records, such as logbook entries, portside samples, survey trawls, etc. (Maunder and Langley 2004; Kristensen et al. 2014; Thorson et al. 2015a). Here, we have instead followed the more conventional approach of using design- or model-based methods to aggregate records across space, where resulting estimates of total catch, abundance, or composition data can then be fitted in a separate stock assessment model. Both the former (fitting assessments directly to observation records) and the latter (analysing records and then fitting in a

Fig. 5. Prediction of proportion-at-length $\hat{P}_c(t)$ for length classes of lingcod in the US portion of the California Current using a design-based (blue) or spatiotemporal model-based (red) estimator including the predicted sample size for each (see Fig. 1 caption for details; length classes are defined as "[X-Y]", representing all fish with length $X \leq l \leq Y$), showing females (*a*) and males (*b*) separately and where any length bin not shown has a proportion of 0.0 for all years. [Colour online.]

(a)
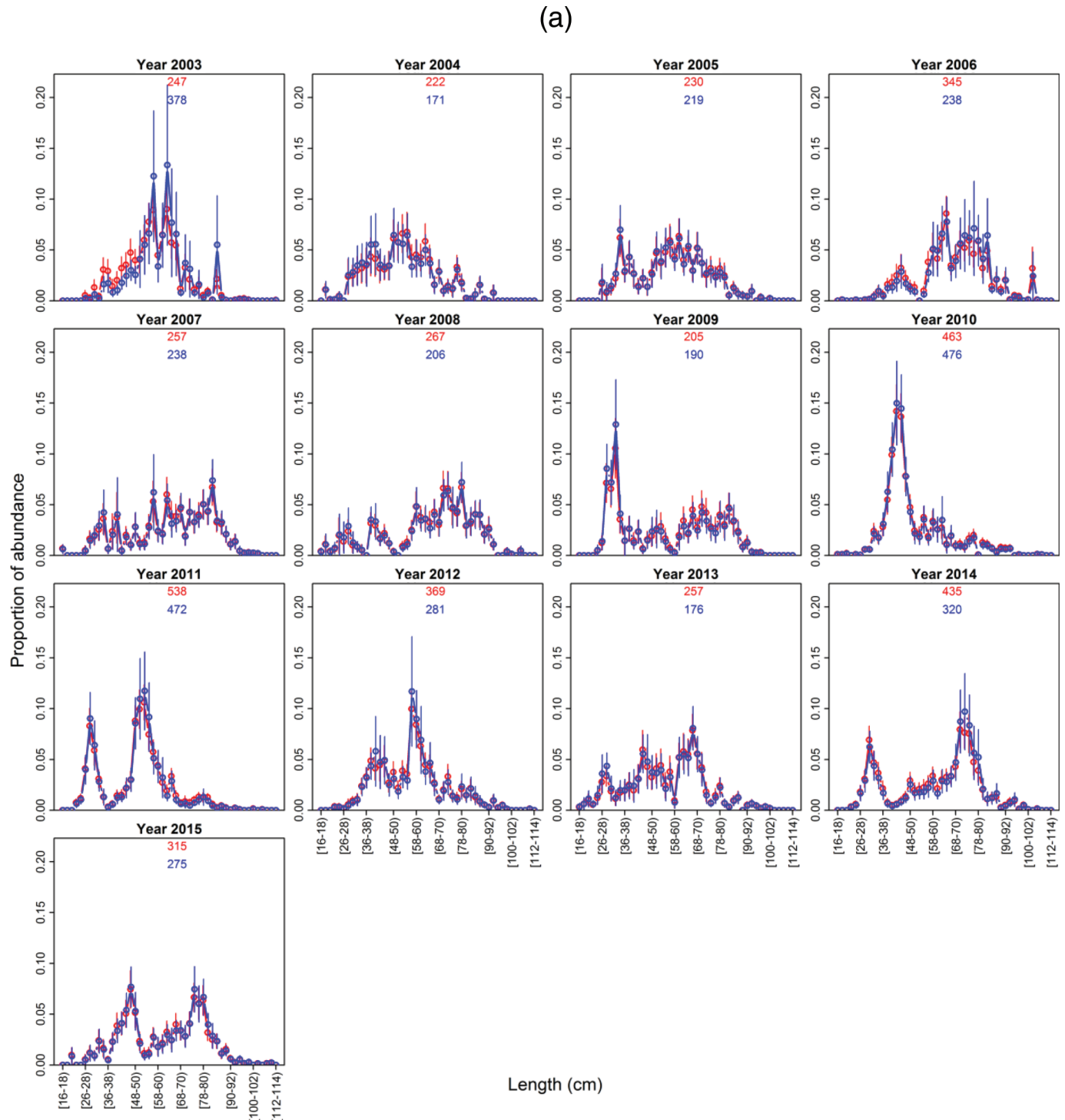
separate assessment model) have strengths and weaknesses. Fitting assessments directly to observation records is likely to have large issues with data weighting (i.e., where observation records are thought to have more information than is appropriate), making estimates of stock status implausibly precise (e.g., Maunder and Langley 2004). However, this over-weighting may be resolved by incorporating new forms of "process error" (e.g., spatiotempo-

ral variation in fish density; Kristensen et al. 2014; Thorson et al. 2015a). By contrast, analysing records and then fitting a secondary assessment model has several advantages, including (*i*) faster parameter estimation and therefore greater scope for biological detail in population dynamics (e.g., Taylor and Methot 2013) and (*ii*) easier statistical diagnostics and explanation of model structure. We note that future research could modify the multivariate
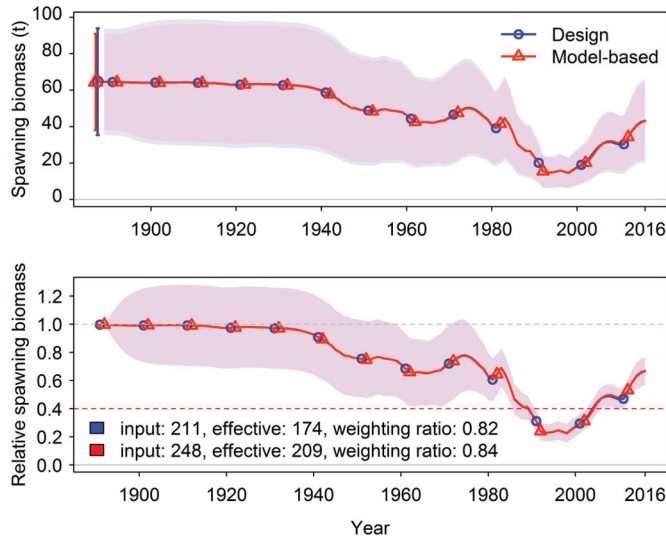
Fig. 5 (*concluded*).

(b)
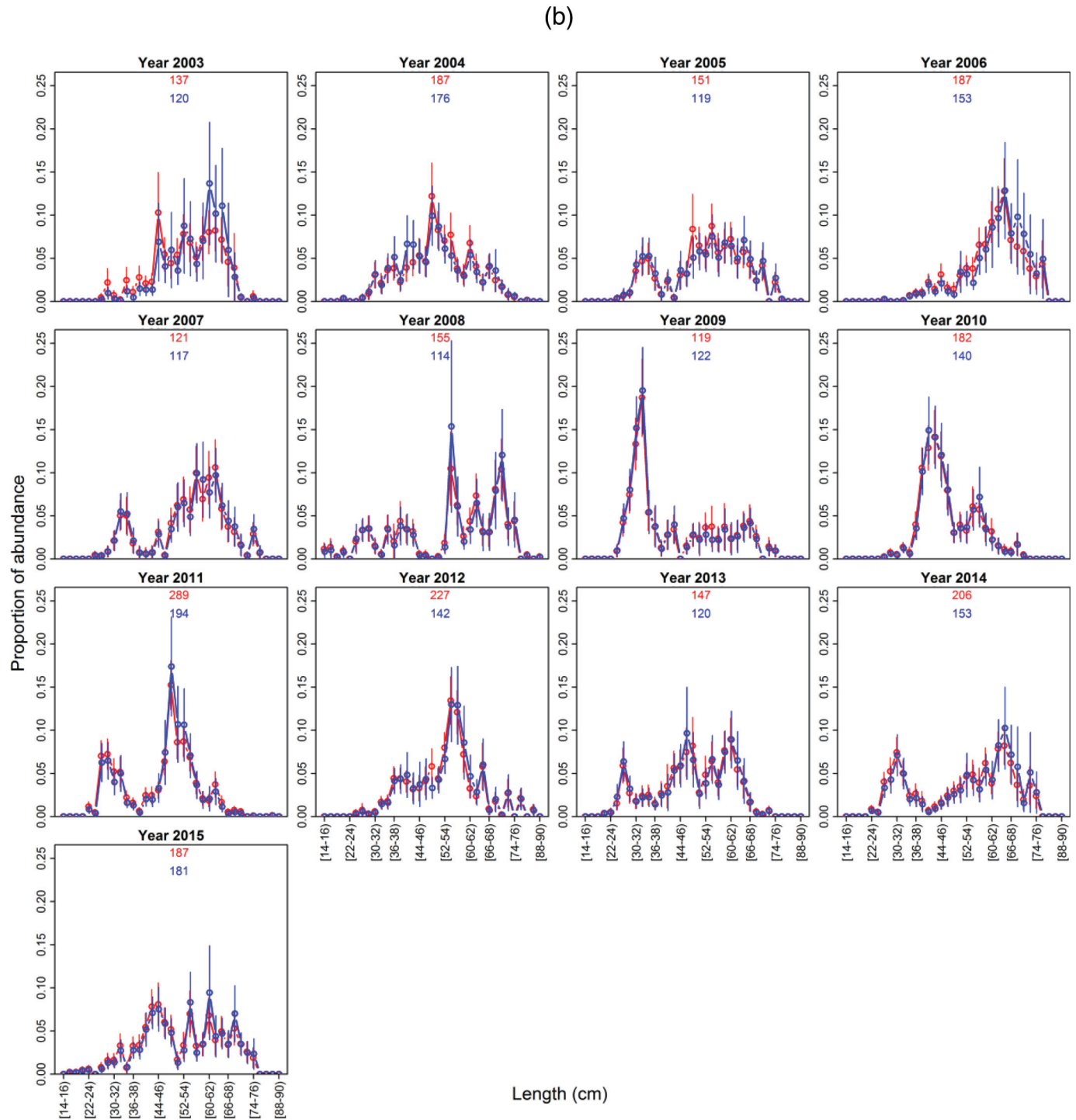
spatiotemporal model used here to incorporate individual growth and fishing mortality, and this would then be a spatiotemporal size-structured population model. We therefore recommend on-going research regarding the practical advantages and drawbacks to these two general approaches to stock assessment models.

We recommend further research comparing spatiotemporal and design-based approaches with compositional expansion using real-world stock assessments. Although spatiotemporal estimates of abundance indices can also be influential in some case studies (e.g., Cao et al. 2017), the model-based abundance indices have not made a large difference in many other stock assessments

(e.g., Thorson and Wetzel 2015). The small impact of changes to abundance indices in some stock assessments is not surprising, given that many assessments are driven primarily by compositional data (Francis 2011). We therefore hypothesize that the impact of using design- versus model-based approaches on stock assessment results will be greater when estimating compositional data than abundance indices.

Previous studies have also used spatiotemporal models to estimate abundance indices for individual sizes or ages (Thorson et al. 2017c; Berg et al. 2014; Kai et al. 2017). In particular, Berg et al. (2014) used spatiotemporal models for estimating age-specific

412

Can. J. Fish. Aquat. Sci. Vol. 76, 2019

Fig. 6. Estimated spawning biomass (top panel, in metric tonnes) and spawning output relative to average unfished levels (bottom panel) showing the management target (line) and 95% confidence intervals (shading) using either design-based (blue) or spatiotemporal model-based (red) estimates of proportion-at-length $\hat{P}_c(t)$ for the NWFSC shelf-slope survey (purple again shows overlap between model- and design-based estimators). We also show the average input sample size, $\tau(t)$, the average effective sample size estimated by the Dirichlet-multinomial approach, and the ratio of these two sample sizes where this "weighting ratio" measures the extent to which compositional data are down-weighted due to lack of fit when combined with other data sources and model assumptions (bottom plot; blue: design-based; red: model-based). [Colour online.]

Fig. 7. Schematic showing role of model-based expansion of compositional data when generating an input sample size $\hat{\tau}(t)$ and estimated composition data $\hat{P}_c(t)$, which are combined with other expanded data $\hat{D}$ in a stock assessment model, where the match between $\hat{P}_c(t)$ and predicted proportions $\pi_c(t)$ depends upon model mis-specification and where any overdispersion results in an estimated effective sample size $\hat{\tau}_{\text{eff}}(t) < \hat{\tau}(t)$. This decreased effective sample size can then be used as a diagnostic to justify added complexity within the stock assessment model, which will affect the degree of model mis-specification and where this feedback between $\hat{\tau}_{\text{eff}}(t)$ and added model complexity (shown in the grey box) can be iterated until the model has satisfactory fit to composition data.
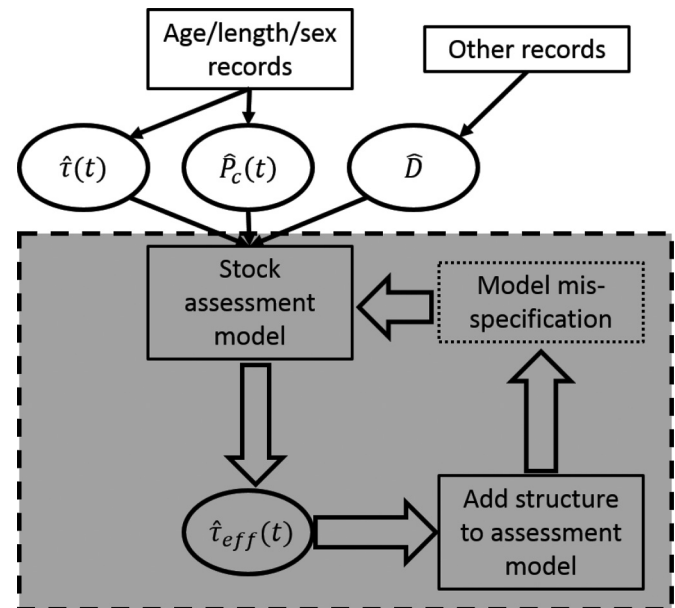

abundance indices for use in a stock assessment model. Berg et al. (2014) concluded that (i) spatiotemporal models were more precise than a design-based estimator, and (ii) estimated abundance indices were correlated for adjacent ages. Including a size-specific abundance index in a stock assessment model is common in Europe and the Northeast US. By contrast, we have estimated a separate time series of abundance index and compositional data, following common practice in the US West Coast. The practice of estimating abundance indices for each size or age is convenient, because the estimated imprecision could be included as a covariance matrix (to account for estimation covariance). By contrast, our model structure has an independent intercept for each size and year, which is designed to minimize the estimation covariance for size composition between sizes and years. We recommend that future research be conducted to explore the impact of including estimation covariance for size-based abundance indices (similar to Berg et al. 2014) or using model specifications that minimize estimation covariance (as we do here) to determine which approach has better performance in existing stock assessments. However, any attempt to include estimation covariance will require changing the input format for existing stock assessment models (e.g., Stock Synthesis for our lingcod case study) to allow the assessment software to read in a covariance matrix associated with every existing row of age or length composition data. We therefore believe this comparison requires a separate research effort.

Finally, we recommend several lines of research to explore the proposed spatiotemporal model for composition data:

1. Data weighting. Future research should explore the impact of model-based data processing on data weighting in stock assessments. To do so, research could simulate a fishery for a size-structured spatiotemporal population, spatial records of

catch rate and compositional samples, and conduct design- or model-based expansion of records. It could then use these estimates to simulate assessment models while estimating overdispersion and then add process errors to account for unmodeled spatial processes implicitly. This simulation experiment would illustrate the potential for estimating input sample size, overdispersion, and process errors to resolve ongoing issues with data weighting in stock assessment. We hypothesize that improved precision when using a spatiotemporal model to expand compositional data would in turn improve estimates of time-varying selectivity.

2. Conditional age-at-length (CAAL) data. Future research could also explore potential improvements in precision when using a spatiotemporal model to expand CAAL data. CAAL data are typically included in Stock Synthesis models to both (i) account for a nonrandom sampling design for individuals that have their ages read and (ii) allow estimating variation in size-at-age among individuals. However, CAAL will require modelling many length bins for each observed age, and this will be computationally challenging for long-lived species. We imagine that future efforts could separately model CAAL data for each survey year, and this will therefore reduce computational costs for each individual year. However, we do not currently know whether a spatiotemporal approach to CAAL data will be computationally feasible.

3. Reviewer standards and assessment terms of reference. Improvements in stock-assessment methods must be accompanied by improvements in how to specify and review stock assessments. We therefore recommend ongoing research regarding diagnostics for spatiotemporal models (including those used here), as well as simulation research to develop guide-

lines for when to use (or not use) model-based methods for compositional data. Model-based methods for abundance indices were developed in part to deal with variation in catch rates among contracted fishing vessels (Helser et al. 2004; Thorson and Ward 2014), and future research could start by showing (*i*) whether vessels also differ in size selectivity and (*ii*) whether this impacts estimated sample sizes and proportions.

These three research lines will require substantial research over the coming years. However, the lingcod case study used here highlights the potential increase in stock-assessment precision arising from spatiotemporal approaches to compositional data. We therefore hypothesize that improvements in compositional analysis will be worth the necessary research effort.

Acknowledgements

References

Albertsen, C.M., Nielsen, A., and Thygesen, U.H. 2016. Choosing the observational likelihood in state-space stock assessment models. Can. J. Fish. Aquat. Sci. **74**(5): 779–789. doi:10.1139/cjfas-2015-0532.

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. Fish. Res. **151**: 91–99. doi:10.1016/j.fishres.2013.10.005.

Binion-Rock, S.M., Reich, B.J., and Buckel, J.A. 2019. A spatial kernel density method to estimate diet composition of fish. Can. J. Fish. Aquat. Sci. **76**(2). [Online ahead of print.] doi:10.1139/cjfas-2017-0306.

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. Can. J. Fish. Aquat. Sci. **74**(11): 1781–1793. doi:10.1139/cjfas-2016-0137.

Cotter, A., and Pilling, G. 2007. Landings, logbooks and observer surveys: improving the protocols for sampling commercial fisheries. Fish Fish. **8**(2): 123–152. doi:10.1111/j.1467-2679.2007.00241.x.

Fortin, M.-J., and Jacquez, G.M. 2000. Randomization tests and spatially autocorrelated data. Bull. Ecol. Soc. Am. **81**(3): 201–205.

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. **68**(6): 1124–1138. doi:10.1139/f2011-025.

Haltuch, M.A., Wallace, J., Akselrud, C.A., Nowlis, J., Barnett, L.A.K., Valero, J.L., et al. 2017. 2017 Lingcod stock assessment [online]. Pacific Fisheries Management Council, Portland, Ore. Available from http://www.pcouncil.org/groundfish/stock-assessments/.

Helser, T.E., Punt, A.E., and Methot, R.D. 2004. A generalized linear mixed model analysis of a multi-vessel fishery resource survey. Fish. Res. **70**(2–3): 251–264. doi:10.1016/j.fishres.2004.08.007.

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. Can. J. Fish. Aquat. Sci. doi:10.1139/cjfas-2016-0327.

Kass, R.E., and Steffey, D. 1989. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). J. Am. Stat. Assoc. **84**(407): 717–726. doi:10.2307/2289653.

Keller, A.A., Wallace, J.R., and Methot, R.D. 2017. The Northwest Fisheries Science Center's west coast groundfish bottom trawl survey: history, design, and description. NOAA Technical Memorandum, Northwest Fisheries Science Center, Seattle, Wash.

Kristensen, K., Thygesen, U.H., Andersen, K.H., and Beyer, J.E. 2014. Estimating spatio-temporal dynamics of size-structured populations. Can. J. Fish. Aquat. Sci. **71**(2): 326–336. doi:10.1139/cjfas-2013-0151.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B.M. 2016. TMB: Automatic differentiation and Laplace approximation. J. Stat. Softw. **70**(5): 1–21. doi:10.18637/jss.v070.i05.

Lindgren, F., Rue, H., and Lindström, J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. **73**(4): 423–498. doi:10.1111/j.1467-9868.2011.00777.x.

Lunsford, C.R., Hulson, P.-J.F., Shotwell, S.K., and Hanselman, D.H. 2015. 12. Assessment of the Dusky Rockfish stock in the Gulf of Alaska. *In* Stock assessment and fishery evaluation report for the groundfish fisheries of the Gulf of Alaska [online]. North Pacific Fishery Management Council, Seattle, Wash. pp. 1013–1102. Available from http://www.afsc.noaa.gov/refm/stocks/plan_team/GOAdusky.pdf.

Maunder, M.N. 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: Estimating the effective sample size. Fish. Res. **109**(2): 311–319. doi:10.1016/j.fishres.2011.02.018.

Maunder, M.N., and Langley, A.D. 2004. Integrating the standardization of catch-per-unit-of-effort into stock assessment models: testing a population dynamics model and using multiple data types. Fish. Res. **70**(2): 389–395. doi:10.1016/j.fishres.2004.08.015.

Maunder, M.N., and Punt, A.E. 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. **142**: 61–74. doi:10.1016/j.fishres.2012.07.025.

Maunder, M.N., Crone, P.R., Punt, A.E., Valero, J.L., and Semmens, B.X. 2017. Data conflict and weighting, likelihood functions and process error. Fish. Res. **192**: 1–4. doi:10.1016/j.fishres.2017.03.006.

Methot, R.D., and Wetzel, C.R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. **142**: 86–99. doi:10.1016/j.fishres.2012.10.012.

Monnahan, C.C., Ono, K., Anderson, S.C., Rudd, M.B., Hicks, A.C., Hurtado-Ferro, F., et al. 2016. The effect of length bin width on growth estimation in integrated age-structured stock assessments. Fish. Res. **180**: 103–112. doi:10.1016/j.fishres.2015.11.002.

Nielsen, A., and Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fish. Res. **158**: 96–101. doi:10.1016/j.fishres.2014.01.014.

Okamura, H., Morita, S.H., Funamoto, T., Ichinokawa, M., and Eguchi, S. 2017. Target-based catch-per-unit-effort standardization in multispecies fisheries. Can. J. Fish. Aquat. Sci. **75**(3): 452–463. doi:10.1139/cjfas-2016-0460.

R Core Team. 2016. R: A language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/.

Sampson, D.B., and Scott, R.D. 2012. An exploration of the shapes and stability of population–selection curves. Fish Fish. **13**(1): 89–104. doi:10.1111/j.1467-2979.2011.00417.x.

Schlather, M. 2009. RandomFields: Simulation and Analysis of Random Fields [online]. Available from http://CRAN.R-project.org/package=RandomFields.

Shelton, A.O., Dick, E.J., Pearson, D.E., Ralston, S., Mangel, M., and Walters, C. 2012. Estimating species composition and quantifying uncertainty in multispecies fisheries: hierarchical Bayesian models for stratified sampling protocols with missing data. Can. J. Fish. Aquat. Sci. **69**(2): 231–246. doi:10.1139/f2011-152.

Skaug, H., and Fournier, D. 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. Comput. Stat. Data Anal. **51**(2): 699–709. doi:10.1016/j.csda.2006.03.005.

Stauffer, G. 2004. NOAA protocols for groundfish bottom trawl surveys of the nation's fishery resources [online]. NOAA Technical Memorandum, National Oceanic and Atmospheric Administration (NOAA), Seattle, Wash. Available from http://www.mafmc.org/s/NOAA-protocols-for-bottom-trawl-surveys.pdf.

Stephens, A., and MacCall, A. 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fish. Res. **70**(2): 299–310. doi:10.1016/j.fishres.2004.08.009.

Stewart, I.J., and Hamel, O.S. 2014. Bootstrapping of sample sizes for length- or age-composition data used in stock assessments. Can. J. Fish. Aquat. Sci. **71**(4): 581–588. doi:10.1139/cjfas-2013-0289.

Taylor, I.G., and Methot, R.D. 2013. Hiding or dead? A computationally efficient model of selective fisheries mortality. Fish. Res. **142**: 75–85. doi:10.1016/j.fishres.2012.08.021.

Thorson, J.T. 2014. Standardizing compositional data for stock assessment. ICES J. Mar. Sci. J. Cons. **71**(5): 1117–1128. doi:10.1093/icesjms/fst224.

Thorson, J.T. 2018. Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. Can. J. Fish. Aquat. Sci. **75**(9): 1369–1382. doi:10.1139/cjfas-2017-0266.

Thorson, J.T., and Barnett, L.A.K. 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES J. Mar. Sci. **74**(5): 1311–1321. doi:10.1093/icesjms/fsw193.

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. Fish. Res. **175**: 66–74. doi:10.1016/j.fishres.2015.11.016.

Thorson, J.T., and Ward, E.J. 2014. Accounting for vessel effects when standardizing catch rates from cooperative surveys. Fish. Res. **155**: 168–176. doi:10.1016/j.fishres.2014.02.036.

Thorson, J.T., and Wetzel, C.R. 2015. The status of canary rockfish (*Sebastes pinniger*) in the California Current in 2015. Pac. Fish. Manag. Counc. **7700**: 97200–1384.

Thorson, J.T., Stewart, I.J., and Punt, A.E. 2012. Development and application of an agent-based model to evaluate methods for estimating relative abundance indices for shoaling fish such as Pacific rockfish (*Sebastes* spp.). ICES J. Mar. Sci. **69**(4): 635–647. doi:10.1093/icesjms/fss003.

Thorson, J.T., Ianelli, J.N., Munch, S.B., Ono, K., and Spencer, P.D. 2015a. Spatial delay-difference models for estimating spatiotemporal variation in juvenile production and population abundance. Can. J. Fish. Aquat. Sci. **72**(12): 1897–1915. doi:10.1139/cjfas-2014-0543.

414

Can. J. Fish. Aquat. Sci. Vol. 76, 2019

Thorson, J.T., Shelton, A.O., Ward, E.J., and Skaug, H.J. 2015b. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES J. Mar. Sci. J. Cons. **72**(5): 1297–1310. doi:10.1093/icesjms/fsu243.

Thorson, J.T., Fonner, R., Haltuch, M.A., Ono, K., and Winker, H. 2017a. Accounting for spatiotemporal variation and fisher targeting when estimating abundance from multispecies fishery data. Can. J. Fish. Aquat. Sci. **74**(11): 1794–1807. doi:10.1139/cjfas-2015-0598.

Thorson, J.T., Johnson, K.F., Methot, R.D., and Taylor, I.G. 2017b. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. Fish. Res. **192**: 84–93. doi:10.1016/j.fishres.2016.06.005.

Thorson, J.T., Ianelli, J.N., and Kotwicki, S. 2017c. The relative influence of temperature and size-structure on fish distribution shifts: A case-study on Walleye pollock in the Bering Sea. Fish Fish. **18**(6): 1073–1084. doi:10.1111/faf.12225.

Winker, H., Kerwath, S.E., and Attwood, C.G. 2013. Comparison of two approaches to standardize catch-per-unit-effort for targeting behaviour in a multispecies hand-line fishery. Fish. Res. **139**: 118–131. doi:10.1016/j.fishres.2012.10.014.

Xu, H., Thorson, J.T., Taylor, I.G., and Methot, R. 2019. A new semiparametric method for autocorrelated age- and time-varying selectivity in age-structured assessment models. Can. J. Fish. Aquat. Sci. **76**(2). [Online ahead of print.] doi:10.1139/cjfas-2017-0446.

Ye, Y., and Dennis, D. 2009. How reliable are the abundance indices derived from commercial catch-effort standardization? Can. J. Fish. Aquat. Sci. **66**(7): 1169–1178. doi:10.1139/F09-070.