Fisheries Research 147 (2013) 426–433

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Accounting for space–time interactions in index standardization models

James T. Thorson[a,*], Eric J. Ward[b]

[a] Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Boulevard East, Seattle, WA 98112, United States
[b] Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Boulevard East, Seattle, WA 98112, United States

ARTICLE INFO

Article history:
Received 21 September 2012
Received in revised form 15 March 2013
Accepted 18 March 2013

Keywords:
Index standardization
Mixed-effects model
Bayesian model
Random effect

ABSTRACT

Scientific survey data are used to estimate abundance trends for fish populations worldwide, and are frequently analyzed using delta-generalized linear mixed models (delta-GLMMs). Delta-GLMMs incorporate information about both the probability of catch being non-zero (catch probability) and the expected value for non-zero catches (catch rates). Delta-GLMMs generally incorporate year as a main effect, and frequently account for spatial strata and/or covariates. Many existing delta-GLMMs do not account for random or systematic differences in catch probability or rates in particular combinations of spatial strata and year (i.e., space–time interactions), and do not recognize potential correlation in random space–time interactions between catch probability and catch rates. We therefore develop a Bayesian delta-GLMM that estimates correlations between catch probability and rates, and compare it with either (a) ignoring year–strata interactions, (b) modeling year–strata interactions as fixed effects, or (c) estimating year–strata interactions in catch probability or rates as independent random effects. These four models are fitted to bottom trawl survey data for 28 species off the U.S. West Coast. The posterior median of the correlation is positive for the majority (18) of species, including all five for which the posterior distribution has little overlap with zero. However, estimating this correlation has little impact on resulting abundance indices or credible intervals. We therefore conclude that the correlated random model will have a little impact on index standardization of the West Coast bottom trawl dataset. However, we propose that the correlated model can quickly identify correlations between occupancy probability and density, and provide our code to allow researchers to quickly identify whether such a correlation is likely to be significantly different from zero for their chosen data set.

Published by Elsevier B.V.

1. Introduction

Scientific surveys of marine populations are conducted worldwide and are an important source of information about abundance trends in marine species. Indices of annual stock abundance obtained from survey data are generally incorporated into population dynamics models to estimate stock productivity, current stock status, and allowable catch levels (Quinn and Deriso, 1999). These outputs are then used to inform fisheries management policy such as annual catch and/or fishing effort levels.

Scientific survey data usually include catch and effort statistics for each survey occasion, and catch per unit effort (CPUE) is commonly treated as a measure of local densities. CPUE data can be summarized using simple statistics (i.e., mean CPUE in a design-based estimator) or complex standardization models to provide an index of stock abundance. Index standardization models provide several benefits over simple summary statistics when analyzing survey data, including: (1) improving estimates of indices and confidence/credible intervals (Ye and Dennis, 2009); (2) estimating the effects of and controlling for variables such as sampling vessel, season, and spatial location; (3) incorporating auxiliary information such as fishing hook depth (Bigelow and Maunder, 2007); (4) accounting for stratified and unbalanced sampling designs, i.e., where particular areas and/or years are sampled more or less heavily than others due to intentional sampling design or random chance (Wiedenmann and Essington, 2006); and (5) accounting for a highly non-normal catch rate distribution, e.g., as caused by fish schooling behaviors (Thorson et al., 2012, 2011).

Delta-generalized linear models (delta-GLMs) have become widely used in fisheries and other fields for index standardization because they allow separation of the model into two biologically

* Corresponding author. Tel.: +1 206 302 1772; fax: +1 206 860 3217.
E-mail addresses: James.Thorson@noaa.gov, James.T.Thorson@gmail.com (J.T. Thorson).

0165-7836/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.fishres.2013.03.012

meaningful components (Pennington, 1983; Stefansson, 1996). The first component uses the detection or non-detection of the species at each survey location to estimate the probability of encountering the target species. The proportion of encounters can change over time due to changes in stock range and abundance, or alternatively this model component may be useful to account for overdispersion. The second component uses the catch rate for survey occasions where the species was detected to estimate population density within its occupied range. Overall stock abundance is then calculated as the product of stock range and densities within the occupied range. Both model components are necessary because marine species will often undergo changes in densities and stock range over time. Delta-GLMs approximate the expected value of each model component as a function of covariates, and have been extended into a mixed-effects modeling framework (delta-GLMMs), allowing model coefficients to also be treated as random effects that differ among grouping variables (e.g., year and strata, Helser et al., 2004). Delta-GLMMs used for index standardization always include a coefficient representing 'Year' for both presence/absence and positive catch model components because the focus of inference during index standardization is on changes in abundance from one year to the next. Standardization models also frequently include 'Area' (a factor representing different sampling regions or strata), for example, to account for consistent differences in stock density between on- and off-shore habitats (Maunder and Punt, 2004).

An interaction between year and area (Year × Area) can be included in the delta-GLMM to represent abundance changes that differ among spatial areas in a random or systematic manner. Systematic changes in abundance among areas, e.g., caused by different fishery exploitation rates in different portions of the stock range, are treated by including Year × Area by estimating a separate parameter representing stock abundance in each year and area (i.e., as a fixed effect). Random differences, e.g., caused by random annual changes in the distribution of total stock abundance among multiple regions, are treated by including Year × Area as a random effect. Treating Year × Area as a random effect requires estimating an additional coefficient (the variance of random Year × Area deviations), but is sometimes more parsimonious than treating Year × Area as a fixed effect because random coefficients are 'shrunk' toward zero (Gelman and Hill, 2007). Random effects are appropriate when each random effect coefficient is exchangeable (e.g., believed to arise from a 'random,' independent, and identical distribution) for each year–area combination (Bolker et al., 2009). Treating Year × Area as random will explain some portion of residual variance, likely resulting in tighter estimates of index credible intervals and more precise index estimates.

Delta-GLMMs generally treat random effects regarding the probability of non-zero catch and the expected catch for non-zero catches in a given year–area combination as statistically independent. This assumption may commonly be violated. As one example, random environmental changes may cause a population to move into different spatial strata than usual. For strata into which they move, the probability of non-zero catch and the expected catch for non-zero catches will both be increasing, causing a positive correlation between Year × Area coefficients. By contrast, negative correlations may also occur due to environmental conditions that cause the target species to be more tightly aggregated within a given year and area. This negative correlation could occur for example with Peruvian anchoveta (*Engraulis ringens*), which aggregates near shore during El Nino events (Gutiérrez et al., 2007), in turn leading to increased densities and decreased encounter rates in any strata that includes both inshore and offshore areas. In this case, the random effect for the catch rate component will be negatively correlated with the random effect for presence/absence component, in violation of the common delta-GLMM assumption that all random effects are exchangeable. This violation of random-effect model assumptions could be rectified by estimating a correlation between random effects within a given year–area combination, although we know of no fisheries studies or models in other fields that have explored this.

In this manuscript, we pursue two related study objectives. First, we determine whether a correlation between model components is supported by available data for U.S. West Coast species and, if so, whether it is generally positive or negative. Second, we explore whether introducing a correlation between model components affects estimated indices of abundance. To accomplish these objectives, we compare four alternative treatments of spatial and temporal interactions ('space–time interactions') in delta-GLMM models: (1) not including any Year × Area interaction; (2) treating Year × Area as fixed effects separately in both the presence/absence and catch rate model; (3) treating Year × Area as independent random effects separately in both the presence/absence and catch rate model; and (4) treating Year × Area as random effects that are correlated between presence/absence and catch rate model components. We first apply all four models to bottom trawl survey data for 28 species obtained from the Northwest Fisheries Science Center (NWFSC) shelf and slope survey conducted annually by the National Oceanic and Atmospheric Administration (NOAA) off the U.S. West Coast (e.g., 2003–2011, Bradburn et al., 2011). We then assess the direction and magnitude of random effect correlations, to identify whether correlations between presence/absence and positive catch components are likely to be positive, negative, or indistinguishable from zero for these 28 species (objective 1). Next, we compare resulting abundance indices and credible interval widths for each model to determine whether including the correlation parameter affects the estimated abundance indices or their variance estimates (objective 2).

2. Methods

2.1. Data availability

As a case study, we use bottom trawl catch and effort data obtained from the NMFS NWFSC shelf and slope survey off the U.S. West Coast (Bradburn et al., 2011; Keller et al., 2008). This survey has operated with similar sampling protocols since 2003, and uses a stratified random sampling design with six sampling strata composed of three depths categories (55–183 m, 184–549 m, and 550–1280 m) and two latitude categories (32–34.5° and 34.5–50° latitude), wherein sampling intensity is approximately equal to the area in each stratum. For the purposes of this study, we adopted the default post-stratification used by the NWFSC (A. Hicks, pers. comm., 2012), which uses the three depth categories from the sampling design and five latitudinal categories (32–34.5°, 34.5–40.5°, 40.5–43°, 43–47.5°, and 47.5–50° latitude). This results in 15 spatial strata.

We compiled data for 5756 sampling tows from 2003 to 2011 by the NWFSC shelf/slope survey. We analyze data for 28 finfish species (Table 1) that were chosen because they either (a) are flatfishes, (b) were assessed in 2011, (c) were likely to be assessed in 2013, or (d) had 200–500 positive catches between 2003–2008 ('positive catches' representing occasions when the species was detected on a sampling occasion), where this level was chosen to ensure that species had information for estimating both presence/absence and positive catch rate model components. These species range widely in the probability of occurrence from 84.1% (Dover sole) to 1.7% (yelloweye rockfish), and in the average catch for positive tows from 1.6 kg (cowcod) to 78.6 kg (chilipepper rockfish). Within each year, sampling for the NWFSC shelf/slope survey occurs in the same time period (May–October) and represents a snapshot of annual biomass that is comparable among years.

Table 1
List of species, with the proportion of tows that result in positive catch for that species, and the average catch in kilograms for every positive catch.

Common name	Scientific name	Proportion with positive catch	Average positive catch (kg)
Arrowtooth	*Atheresthes stomias*	0.315	24.306
Aurora	*Sebastes aurora*	0.145	5.411
Bocaccio	*Sebastes paucispinis*	0.044	10.398
Canary	*Sebastes pinniger*	0.054	39.471
Chilipepper	*Sebastes goodie*	0.114	78.603
Cowcod	*Sebastes veils*	0.021	1.613
Darkblotched	*Sebastes crameri*	0.167	12.403
Dover	*Microstomus pacificus*	0.841	55.004
English	*Parophrys vetulus*	0.326	10.780
Greenspotted	*Sebastes chlorostictus*	0.050	5.956
Greenstriped	*Sebastes elongates*	0.214	11.159
Halfbanded	*Sebastes semicintus*	0.058	22.722
Hake	*Merluccius productus*	0.530	33.669
Longspine thornyhead	*Sebastolobus alascanus*	0.429	46.886
Petrale	*Eopsetta jordani*	0.333	8.395
Pacific Ocean perch	*Sebastes alutus*	0.069	29.355
Redbanded	*Sebastes babcocki*	0.084	1.828
Rosethorn	*Sebastes helvomaculatus*	0.076	6.738
Rougheye	*Sebastes aleutianus*	0.047	5.272
Sablefish	*Anoplopoma fimbria*	0.712	20.426
Sanddab	*Citharichthys sordidus*	0.245	34.733
Sharpchin	*Sebastes zacentrus*	0.059	70.163
Shortbelly	*Sebastes jordani*	0.070	53.779
Dogfish	*Squalus acanthias*	0.289	42.247
Shortspine thornyhead	*Sebastolobus alascanus*	0.599	11.999
Widow	*Sebastes entomelas*	0.034	11.395
Yelloweye	*Sebastes ruberrimus*	0.017	7.070
Yellowtail	*Sebastes flavidus*	0.050	63.554

2.2. Model overview

We first present the ‘correlated model’ because it is the most complicated of the four models. We then present in turn how each other model is derived from the correlated model. We use a Bayesian hierarchical modeling framework, which specifies prior probabilities for model parameters and the conditional probability of the data given parameters to define the posterior distribution of model parameters. We use Bayesian methods to simplify computation of the bivariate integral used in the correlated model. This integral, while possible to compute in a maximum likelihood statistical framework (e.g., Thorson et al., 2011), is computationally easier using Bayesian Markov chain Monte Carlo methods.

2.2.1. Correlated Stratum × Year effects

The probability of catch C (in weight, i.e., kilograms) being non-zero is approximated by a logistic regression model:

$$p(C>0|s_i,y_i)=\Phi\left(\sum_{i=1}^{n_{\mathrm{strata}}}\omega_j^{(s)}I(s_i=j)+\sum_{i_{\mathrm{year}}=1}^{n_{\mathrm{year}}}\omega_k^{(y)}I(y_i=k)\right.$$
$$\left.+\sum_{i_{\mathrm{strata}}=1}^{n_{\mathrm{strata}}}\sum_{i_{\mathrm{year}}=1}^{n_{\mathrm{year}}}\omega_{j\cdot k}^{(sy)}I(s_i=j)I(y_i=k)\right)\tag{1}$$

where s_i and y_i are strata and year for tow i, $\omega^{(s)}$, $\omega^{(y)}$, and $\omega^{(sy)}$ are parameters representing the effect of stratum, year, and the Stratum × Year interaction on the probability that C is non-zero, n_{strata} and n_{year} are the number of strata and years, respectively, j and k are indices representing strata and year, respectively, Φ is the logistic transformation $\Phi(X)=\exp(X)/(1+\exp(X))$, and $I(x=b)$ is an indicator variable that equals one if $x=b$ and zero otherwise.

The probability density for catch C given that catch is non-zero is approximated by a gamma distribution:

$$p(C=c|C>0,s_i,y_i,a_i)=\mathrm{Gamma}(c|\alpha,\beta_i)\tag{2}$$

where $\alpha=1/CV^2$, $\beta_i=1/(\mu_i\cdot CV^2)$, CV is the estimated coefficient of variation for C given that $C>0$, and μ_i is the expected value of catch for non-zero tow i. We parameterize this gamma distribution such that the CV of all non-zero catches is constant (as represented by the estimated parameter CV). Other distributions may be appropriate (lognormal, inverse Gaussian, etc.) but for skewed distributions that occur in fishery data, the gamma distribution is well behaved (Myers and Pepin, 1990). The μ_i is in turn approximated by an exponential-transformed linear model:

$$\mu_i=a_i\cdot\exp\left(\sum_{j=1}^{n_{\mathrm{strata}}}\gamma_j^{(s)}I(s_i=j)+\sum_{k=1}^{n_{\mathrm{year}}}\gamma_k^{(y)}I(y_i=k)\right.$$
$$\left.+\sum_{j=1}^{n_{\mathrm{strata}}}\sum_{k=1}^{n_{\mathrm{year}}}\gamma_{j,k}^{(sy)}I(s_i=j)I(y_i=k)\right)\tag{3}$$

where a_i is the area swept (in ha) for tow i, and $\gamma^{(s)}$, $\gamma^{(y)}$ and $\gamma^{(sy)}$ are parameters representing the effect of strata, year and the Stratum × Year interaction on the expected value of non-zero catch C_i.

Stratum × Year interactions are treated as random effects, and we specify that the random effect for presence/absence $\omega^{(sy)}$ in strata s and year y is correlated with the random effect for positive catch rates $\gamma^{(sy)}$ in strata s and year y:

$$p(\omega_{j,k}^{(sy)},\gamma_{j,k}^{(sy)}|\Sigma_{sy})=\mathrm{MVN}(0,\Sigma_{sy})\tag{4}$$

where $\mathrm{MVN}(\mu,\Sigma)$ is a multivariate normal density function, and Σ_{sy} is the covariance among Stratum × Year random effects within a given strata–year combination:

$$\Sigma_{sy}=\begin{bmatrix}\sigma_\omega^2 & \rho_{sy}\sigma_\omega\sigma_\gamma\\ \rho_{sy}\sigma_\omega\sigma_\gamma & \sigma_\gamma^2\end{bmatrix}\tag{5}$$

where σ_ω^2 and σ_γ^2 are the variance for positive catch rate and presence/absence random effects, and ρ_{sy} is the estimated correlation between Stratum × Year random effects for presence/absence and positive catch rates model components. This correlated model therefore has a total of 322 estimated coefficients for each species (9 year effects, 15 strata effects, and 135 strata–year effects for each sub-model; 2 strata–year variance parameters, 1 estimated correlation, and 1 dispersion parameter α for the positive model), although the effective degrees of freedom may be less than this number due to random effect shrinkage (Spiegelhalter et al., 2002).

The Bayesian model specification is completed by including prior distributions for all parameters. We used a weakly informative gamma prior on α, $p(\alpha)=\mathrm{Gamma}(0.001,0.001)$, and bounded uniform priors on all fixed-effect parameters $p(X)=1/40$ if $-20<X<20$ and zero otherwise (the value 1/40 ensures that the integral of this prior is one), where X represents $\gamma^{(s)}$, $\gamma^{(y)}$, $\omega^{(s)}$, and $\omega^{(y)}$. Because the correlated strata–year effects are treated as multivariate normal, we used a standard conjugate inverse-Wishart prior on the covariance matrix (Gelman et al., 2003), which has a prior mean of zero (i.e., represents a prior assumption of no correlation). The posterior distribution for the model is then defined as the product of all terms defined previously.

2.2.2. Uncorrelated Year × Area effects

As a second model, we consider allowing Year × Area effects in the presence/absence and positive catch rate components to be independent ('uncorrelated Year × Area') by specifying that correlations are zero ($\rho_{sy} = 0$). We used weakly informative bounded uniform priors on the standard deviation of Stratum × Year random effects $p(X) = 1/20$ if $0 < X < 20$ and zero otherwise, where X represents σ_ω or σ_γ (Gelman, 2006). In this case, Σ_{sy} is diagonal and there remain 321 estimated coefficients for each species.

2.2.3. Fixed Year × Area effects

As a third possible model, we specify that Year × Area effects are estimated as fixed effects. This model eliminates Eqs. (4) and (5), and instead specifies bounded uniform priors on Year × Area effects, $p(X) = 1/40$ if $-20 < X < 20$ and zero otherwise, where X represents $\gamma^{(sy)}$ and $\omega^{(sy)}$. It also requires that all strata and year main effects are set to zero (i.e., $\gamma^{(s)} = \omega^{(s)} = \gamma^{(y)} = \omega^{(y)} = 0$), to prevent confounding of main and interaction effects (where this parameterization additionally minimizes parameter correlations). This leaves 271 estimated coefficients for each species, although the number of effective degrees of freedom is likely higher than for the random Year × Area models due to an absence of shrinkage of interaction coefficients (Spiegelhalter et al., 2002).

2.2.4. Absent Stratum × Year effects

The fourth model we implement involves no estimation of Year × Area effects, but still retains the fixed area and year effects. The model without Year × Area interactions again eliminates Eqs. (4) and (5), and specifies that Year × Area effects are zero, $\gamma^{(sy)} = \omega^{(sy)} = 0$, leaving 47 estimated coefficients per species.

2.3. Estimating an index of abundance

We then use all four models (correlated, uncorrelated, fixed, and absent) to estimate an index of abundance for all 28 species. This index is calculated by multiplying the posterior distributions for the probability of non-zero catch and the probability density of catch when non-zero, and taking the sum weighted by strata area A_j for each stratum j (Table 2):

$$p(B_k|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a}) = \sum_{j}^{n_{\text{strata}}} A_j \cdot p(C_{j,k} > 0|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a}) \cdot p(\mu_{j,k}|C_{j,k} > 0,\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a}) \tag{6}$$

where $p(B_k|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a})$ is the probability density for random variable B_k representing abundance (i.e., biomass) in year k given all available data (i.e., vectors of catch $\boldsymbol{c}$, strata $\boldsymbol{s}$, year $\boldsymbol{y}$, and area swept $\boldsymbol{a}$ for all tows), A_j is total area in stratum j, $p(C_{j,k} > 0|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a})$ is the probability that catch $C_{j,k}$ in strata j and year k is positive, and $p(\mu_{j,k}|C_{j,k} > 0,\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a})$ is the probability density of non-zero catches $\mu_{j,k}$ for that year and stratum:

$$p(C_{j,k} > 0|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a}) = \Phi(p(\omega_j^{(s)}|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a}) + p(\omega_k^{(y)}|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a}) + p(\omega_{j,k}^{(sy)}|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a})) \tag{7}$$

$$p(\mu_{j,k}|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a}) = \exp(p(\gamma_j^{(s)}|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a}) + p(\gamma_k^{(y)}|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a}) + p(\gamma_{j,k}^{(sy)}|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a})) \tag{8}$$

where $p(X|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a})$ is the posterior distribution for parameter X. We use the median of $p(B_k|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a})$ as the index of abundance, and the standard deviation of $p(B_k|\boldsymbol{c},\boldsymbol{s},\boldsymbol{y},\boldsymbol{a})$ divided by its median as a measure of precision for B_k.

Table 2
Comparisons among four candidate models for all 28 species, showing the posterior distribution for the random effect correlation parameter including the median estimate and the posterior probability that the correlation is positive (where bold values indicate that this posterior probability exceeds 0.95).

Species	Median	5%	95%	$\Pr(\rho_{sy} > 0)$
Arrowtooth	−0.016	−0.403	0.373	0.471
Aurora	0.094	−0.381	0.518	0.653
Bocaccio	0.326	−0.343	0.758	0.840
Canary	0.255	−0.317	0.663	0.810
Chilipepper	0.059	−0.452	0.521	0.587
Cowcod	−0.044	−0.700	0.638	0.462
Darkblotched	−0.070	−0.471	0.334	0.374
Dover	0.421	0.154	0.641	**0.999**
English	0.480	0.129	0.726	**0.995**
Greenspotted	0.093	−0.441	0.556	0.626
Greenstriped	−0.043	−0.430	0.354	0.414
Halfbanded	0.001	−0.613	0.662	0.501
Hake	0.307	0.016	0.557	**0.981**
Longspine thornyhead (LST)	−0.075	−0.423	0.293	0.352
Petrale	0.041	−0.350	0.411	0.581
Pacific Ocean perch (POP)	−0.255	−0.682	0.338	0.193
Redbanded	−0.253	−0.684	0.345	0.210
Rosethorn	0.451	−0.068	0.756	**0.960**
Rougheye	−0.089	−0.601	0.526	0.387
Sablefish	0.145	−0.135	0.410	0.842
Sanddab	−0.218	−0.597	0.243	0.177
Sharpchin	0.241	−0.386	0.698	0.772
Shortbelly	0.532	−0.038	0.823	**0.967**
Dogfish	0.079	−0.238	0.376	0.682
Shortspine thornyhead (SST)	0.050	−0.297	0.386	0.611
Widow	0.336	−0.327	0.739	0.851
Yelloweye	0.020	−0.593	0.664	0.522
Yellowtail	−0.008	−0.571	0.558	0.491

2.4. Estimation and convergence

Samples from the posterior distribution are calculated using Markov chain Monte Carlo (MCMC) methods as implemented in JAGS (Plummer, 2003) and called from the R statistical platform (R Development Core Team, 2011) using the R2jags package (Su and Yajima, 2012). We used three chains, each obtaining 70,000 samples, the first 50,000 of which are used to adapt the sampling algorithm and hence are discarded. The remaining 20,000 samples were then thinned to obtain 4000 samples per chain that were approximately independent.

All results that we present are made after checking for evidence of non-convergence. Convergence checks include visual inspection of sampling chains for estimated parameters. We additionally estimate the Gelman–Rubin R statistic, representing the ratio of variance within and between chains, and the first-order autocorrelation for all derived parameters, i.e., the probability of occurrence, density, and resulting abundance for each strata and year. For the total abundance indices across all strata and years, for example, the median Gelman–Rubin R statistic was below 1.1 for nearly all species and models. For two species (cowcod, halfbanded), the median Gelman–Rubin R statistic was less than 1.1 for the fixed effects model, but was in the range 1.2–1.3 for all other models. Similarly, the median absolute autocorrelation across strata and years was <1.03 for all species and the first three models (no strata–year effects, fixed effects, uncorrelated random effects). For the correlated model, which had slower mixing, approximately half of the species had median lag-1 autocorrelations between 0.1 and 0.2.

3. Results

3.1. The direction and importance of correlated random interactions

We first address study objective #1, i.e., whether a correlation between model components is common for U.S. West Coast species and, if so, whether it is generally positive or negative.

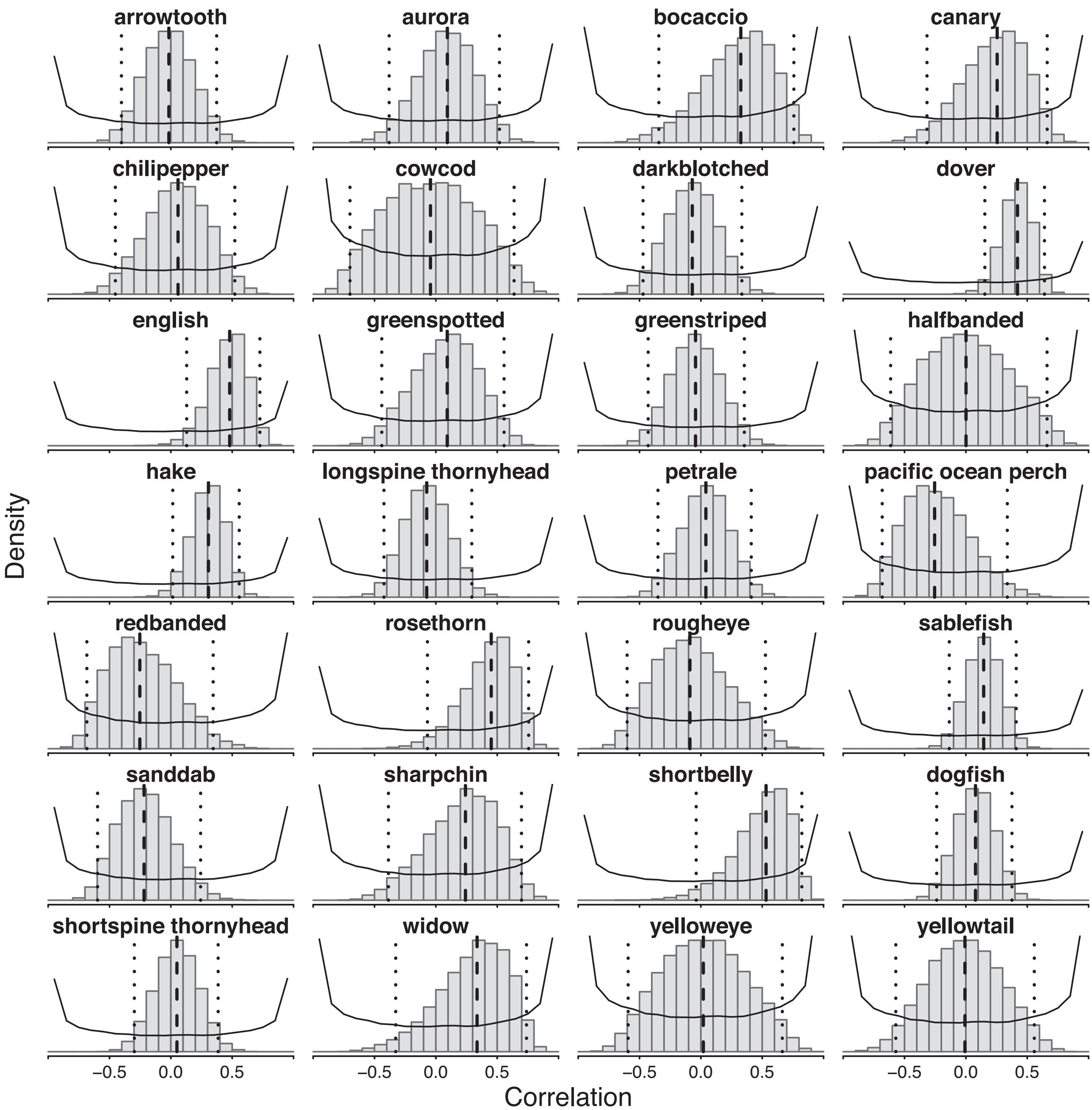

Fig. 1. Posterior distribution for correlation ρ_{sy} between Stratum × Year random effects for positive catch $\gamma^{(sy)}$ and for presence/absence $\omega^{(sy)}$ model components, with median (dashed line) and 95% credible interval (dotted lines) as well as the Wishart prior distribution (solid line).

Specifically, we use the posterior distribution for the random-effect correlation ρ_{sy} to interpret whether the correlation is likely to be non-zero. The posterior for ρ_{sy} is greater than 95% positive for Dover sole, English sole, hake, rosethorn rockfish, and shortbelly rockfish (Table 1). Inspection of the posterior distribution for ρ_{sy} for each species individually (Fig. 1) suggests that other species (e.g., widow) might be positive, although the length of the current dataset a not sufficient to be confident. This positive correlation of Year × Area random effects between presence/absence and positive catches could arise from random changes in spatial distribution among years which causes increased (decreased) abundance in particular strata and years, in turn leading to increased (decreased) probability of detection and catches when detected.

3.2. Comparing abundance index estimates and precision

Next, we address study objective #2, i.e., whether introducing a correlation between model components affects estimated indices of abundance. Comparison of the abundance indices for the random and correlated models for those species where $\Pr[\rho_{sy} > 0] > 0.95$ (Fig. 2) shows that there is little difference in the estimated indices or credible intervals. However, minor exceptions do occur, such as the 10% lower log-standard deviation for the “correlated” model than the “random” model for shortbelly rockfish. Indices for all stocks and models (Supp. Fig. 1) similarly show little difference among index standardization methods. In general, the greatest difference is seen for the model without any Year × Area interactions. This difference arises because the “zero” model is unable to account

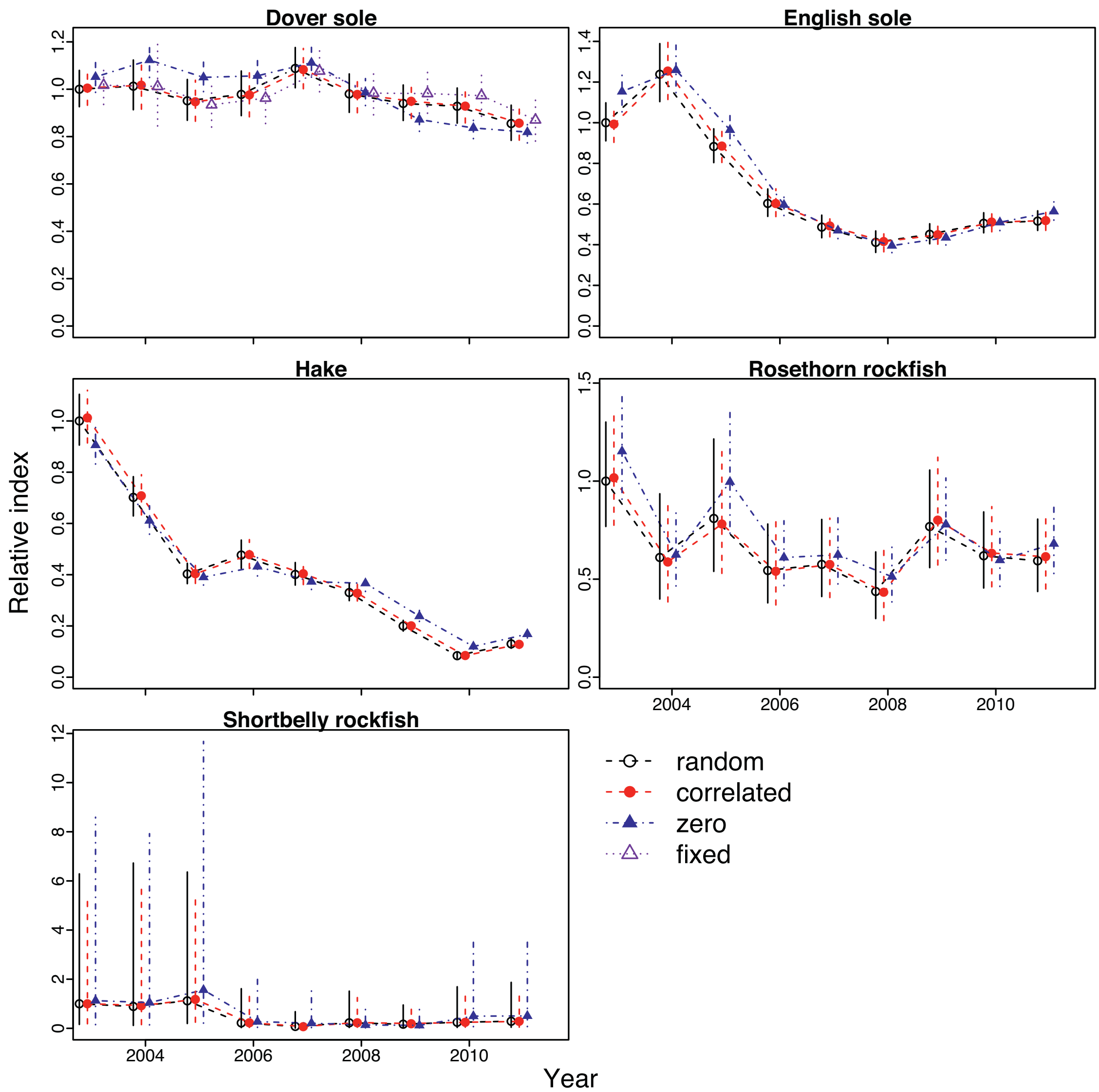

Fig. 2. Indices of abundance (lines, standardized) and ±one posterior standard deviation for those five species where the posterior distribution for correlations was significant ($\Pr[\rho_{sy} > 0] > 0.9$). Lines for are offset along the *x*-axis to improve legibility, and the fixed-effect Stratum × Year model is only included for those species that have at least one encounter in every strata–year combination (e.g., Dover sole).

for any space–time interaction, and hence the difference between the "zero" model and other models is most pronounced for species such as canary rockfish that have previously been shown to be susceptible to extraordinarily large catches (Thorson et al., 2011). We also note a relatively large difference between "fixed" and "zero" models for Dover sole (which has the largest such difference of all 28 species). In this case, the "random" and "correlated" models generate a model that is intermediate between the "fixed" and "zero" models, as is expected given that these random-effect models shrink estimates of strata–year interactions toward zero.

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fishres.2013.03.012.

4. Discussion

Simple survey analysis models for count data imply a linkage between the probability of detection and the expected catch. For example, sampling locations with a greater expected value for non-zero catches will also have a greater probability of obtaining a non-zero catch when using a Poisson-distributed GLM. This correlation between probability of detection and expected catch for count data is also maintained for more-complex models, e.g., zero-inflated negative binomial GLMMs. However, survey analysis models for continuous-valued data (e.g., catch weight per hectare as often used in fisheries stock assessment) have not previously been developed that incorporate a correlation between the probability of positive catch and the expected catch. We rectified this absence by developing a delta-GLMM that estimates a correlation between random effects affecting the positive and presence/absence model components. This model allows for greater comparability with count-data models, and can be used to rapidly evaluate the strength of evidence for such a correlation.

Overall, we find mixed support for a correlation between the probability of positive catch and the expected value of non-zero catch. This correlation has a posterior distribution with little (<5%) overlap with zero for only 5 of 28 species, but is positive for all five of these. A noteworthy caveat is that the strength and magnitude of this correlation could be affected by the spatial strata stratification chosen a priori in any assessment, and the correlation might change if strata were chosen at a more fine or coarse spatial scale. A positive correlation can arise when environmental or anthropogenic factors cause random interannual changes in distribution among spatial strata. Potential environmental factors include transient oceanographic effects, which could cause a relocation of migratory species (e.g., Pacific hake) to suitable habitats (Methot and Dorn, 1994). Potential anthropogenic factors resulting in a correlation include changes over time in the fishing intensity in different strata, which by decreasing total abundance in particular strata will cause a synchronized decrease in the probability of detection and expected catch size in those strata.

For the NWFSC dataset analyzed here, the exact mechanism responsible for observed correlations in Year × Area terms remains unclear. In particular, the species for which correlations are supported differ greatly in their distribution (e.g., Pacific hake is encountered throughout the bottom trawl sampling strata while shortbelly rockfish is primarily encountered in southern California), degree of fishing pressure (shortbelly has close-to-zero fishing mortality, the relative exploitation rate of English sole is close to 2%, and the relative exploitation rate of hake is 10% in recent years; Field et al., 2007; Stewart, 2007; Stewart et al., 2011), migration (rockfish are sedentary, English sole and Pacific hake are migratory) and age at maturity (English sole and Pacific hake have lower age at maturity than most rockfish). As exploratory analysis, we plotted trends in the CPUE-weighted center-of-mass of depth and latitude for each of the 28 species as a proxy for spatial shifts in stock range. While no substantial trends in latitude were apparent, changes in depth were apparent for Dover sole, hake, sablefish, and some rockfish species (rosethorn, widow, and halfbanded; results not shown). Thus, shifts in abundance among depths may explain the positive correlation between the probability of non-zero catch and catch rates for Dover and rosethorn, although this shift in depth distribution could itself be attributed to either environmental or anthropogenic factors.

The correlation between occupancy and expected catch in our analysis is analogous to the occupancy–abundance relationships that have been more broadly studied in macro-ecology (Gaston et al., 2000; Zuckerberg et al., 2008). Positive relationships between occupancy and abundance are well documented, and can arise from a number of causes. Examples include if a species is only surveyed in a portion of its total range (the 'range position' hypothesis), if the spatial distribution of the prey of a target species varies from one year to the next (the 'resource' hypothesis), and if the target species periodically re-colonizes and/or is extirpated from suitable habitat (the 'population dynamics' hypothesis; Gaston et al., 2000). Discriminating between individual mechanisms responsible for producing observed occupancy–abundance relationships is complicated by the fact that each of these hypotheses is capable of generating a range of functional forms for the relationship between occupancy and abundance. For example, Freckleton et al. (2005) illustrated that varying the habitat quality within patches and dispersal between patches could produce occupancy-abundance curves that included nearly flat lines (no relationship), near exponential increases, and asymptotic relationships.

Nevertheless, including the correlation between presence/absence and positive catch rate components had very little effect on either the estimated indices or standard errors for any of these species in our study relative to the random effect model. These species represent a diversity of data qualities (ranging from easy to hard to detect) and life histories (migratory to sedentary). Though we found support for the correlated model in only a handful of cases, we caution that stronger correlations may be found if the strata design was changed (and similarly, some of the correlations we estimated as important may disappear). Each of the species in our analysis has different life history characteristics, ranges, behaviors, and are likely spatially aggregated at different scales. The strata in our analysis were chosen to be representative of general spatial and depth strata, similar to the default strata used in stock assessments for this region. For data rich species, future research will investigate the presence of these correlations across species-specific stratification schemes. We suspect that the correlated model will have little impact on index estimates for other species and study systems with comparable data quantity and quality because, in general, there was little difference in the abundance indices between the random and correlated models. The correlated model may still affect abundance index estimates in study systems with less data within particular strata or years, e.g., surveys before and after establishing no-take areas. In the latter case, continued collection of presence/absence data may be feasible even when the total weight of catch cannot be measured (i.e., because it is not brought onto a sampling vessel).

In summary, we recommend that researchers evaluate multiple treatments of strata–year interactions when analyzing survey data, including the absence of an interaction and its estimation as fixed or random effects. In cases where a random effect models is deemed appropriate, we recommend that researchers additionally explore a model with correlated random effects, and that it be used to estimate indices of abundance for stock assessment purposes whenever the posterior distribution for this correlation differs substantially from zero. However, in light of the small difference in this study between indices estimated by random and correlated models, we do not suggest that adapting the correlated model to other fishery management regions should be a high research priority for stock assessment scientists. Instead, researchers may use our code to rapidly assess whether there is likely to be support for the correlated model for their data sets, and may use this to evaluate the importance of including any such correlation. Our code is provided for this purpose on the Northwest Fisheries Science Center subversion website (https://r-forge.r-project.org/scm/?group_id=1316).

Acknowledgements

We are extremely grateful to the NMFS NWFSC survey team that collects the shelf/slope survey data, including (among many others) J. Buchanan, K. Bosley, and D. Kamikawa. This work was improved by discussions with J. Cope, E.J. Dick, O. Hamel, A. Hicks, O. Shelton, I. Stewart, I. Taylor, and J. Wallace. Data was provided by B. Horness and J. Wallace, sample code was provided by E.J. Dick and J. Wallace, and the manuscript benefited from helpful comments from L. Brooks, J. Hastie, M. McClure, A. Punt, I. Taylor, and two anonymous reviewers.

References

Bigelow, K.A., Maunder, M.N., 2007. Does habitat or depth influence catch rates of pelagic species? Canadian Journal of Fisheries and Aquatic Sciences 64, 1581–1594.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H., White, J.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution 24, 127–135.

Bradburn, M.J., Keller, A.A., Horness, B.H., 2011. The 2003 to 2008 U.S. West Coast Bottom Trawl Szurveys of Groundfish Resources off Washington, Oregon, and California: Estimates of Distribution, Abundance, Length, and Age Composition. Northwest Fisheries Science Center, Seattle, WA.

Field, J.C., Dick, E.J., MacCall, A.D., 2007. Stock Assessment Model for the shortbelly rockfish, *Sebastes jordani*, in the California Current (No. NOAA-TM-NMFS-SWFSC-405). Southwest Fisheries Science Center, Santa Cruz, CA.

Freckleton, R.P., Gill, J.A., Noble, D., Watkinson, A.R., 2005. Large-scale population dynamics, abundance–occupancy relationships and the scaling from local to regional population size. Journal of Animal Ecology 74, 353–364.

Gaston, K.J., Blackburn, T.M., Greenwood, J.J., Gregory, R.D., Quinn, R.M., Lawton, J.H., 2000. Abundance–occupancy relationships. Journal of Applied Ecology 37, 39–59.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1, 515–533.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. Bayesian Data Analysis, second ed. Chapman & Hall, Boca Raton, FL.
Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, Cambridge, UK.
Gutiérrez, M., Swartzman, G., Bertrand, A., Bertrand, S., 2007. Anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*) spatial dynamics and aggregation patterns in the Humboldt Current ecosystem, Peru, from 1983–2003. Fisheries Oceanography 16, 155–168.
Helser, T.E., Punt, A.E., Methot, R.D., 2004. A generalized linear mixed model analysis of a multi-vessel fishery resource survey. Fisheries Research 70, 251–264.
Keller, A.A., Horness, B.H., Fruh, E.L., Simon, V.H., Tuttle, V.J., Bosley, K.L., Buchanan, J.C., Kamikawa, D.J., Wallace, J.R., 2008. The 2005 U.S. West Coast Bottom Trawl Survey of Groundfish Resources off Washington, Oregon, and California: Estimates of Distribution, Abundance, and Length Composition. Northwest Fisheries Science Center, Seattle, WA.
Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research 70, 141–159.
Methot, R.D., Dorn, M.W., 1994. Biology and fisheries of North Pacific hake (*M. productus*). In: Alheit, J., Pitcher, T.J. (Eds.), Hake Biology, Fisheries and Markets, Fish & Fisheries Series. Chapman & Hall, London, pp. 389–414.
Myers, R.A., Pepin, P., 1990. The robustness of lognormal-based estimators of abundance. Biometrics 46, 1185–1192.
Pennington, M., 1983. Efficient estimators of abundance, for fish and plankton surveys. Biometrics 39, 281–286.
Plummer, M., 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March, pp. 20–22.
Quinn, T.J., Deriso, R.B., 1999. Quantitative Fish Dynamics. Oxford University Press, Oxford, UK.
R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. Vienna, Austria.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Linde, A. van der, 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society – Series B (Statistical Methodology) 64, 583–639.
Stefansson, G., 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES Journal of Marine Science 53, 577–588.
Stewart, I.J., 2007. Updated US English Sole Stock Assessment: Status of the Resource in 2007. Northwest Fisheries Science Center, Seattle, WA.
Stewart, I.J., Forrest, R.E., Grandin, C., Hamel, O.S., Hicks, A.C., Martell, S.J.D., Taylor, I.G., 2011. Status of the Pacific Hake (whiting) Stock in US and Canadian Waters in 2011. Northwest Fisheries Science Center, Seattle, WA.
Su, Y.-S., Yajima, M., 2012. R2jags: A Package for Running jags from R.
Thorson, J.T., Stewart, I., Punt, A., 2011. Accounting for fish shoals in single- and multi-species survey data using mixture distribution models. Canadian Journal of Fisheries and Aquatic Sciences 68, 1681–1693.
Thorson, J.T., Stewart, I.J., 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 Journal of Marine Science 69, 635–647.
Wiedenmann, J., Essington, T., 2006. Density-dependent overwinter survival in young-of-year bluefish (*Pomatomus saltatrix*)? A new approach for assessing stage-structured survival. Canadian Journal of Fisheries and Aquatic Sciences 63, 1934–1943.
Ye, Y., Dennis, D., 2009. How reliable are the abundance indices derived from commercial catch-effort standardization? Canadian Journal of Fisheries and Aquatic Sciences 66, 1169–1178.
Zuckerberg, B., Porter, W.F., Corwin, K., 2008. The consistency and stability of abundance–occupancy relationships in large-scale population dynamics. Journal of Animal Ecology 78, 172–181.