ICES Journal of Marine Science

ICES Journal of Marine Science (2015), 72(5), 1297–1310. doi:10.1093/icesjms/fsu243

Original Article

Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes

James T. Thorson[1]*, Andrew O. Shelton[2], Eric J. Ward[2], and Hans J. Skaug[3]

[1]*Fisheries Resource Assessment and Monitoring Division (FRAM), Northwest Fisheries Science Center, National Marine Fisheries Service (NMFS), NOAA, 2725 Montlake Boulevard E, Seattle, WA 98112, USA*

[2]*Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service (NMFS), NOAA, 2725 Montlake Boulevard E, Seattle, WA 98112, USA*

[3]*Department of Mathematics, University of Bergen, PO Box 7800 5020 Bergen, Norway*

**Corresponding author: tel: +1 206 302 1772; fax: +1 206 860 6792; e-mail: James.Thorson@noaa.gov*

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

Received 25 August 2014; revised 18 November 2014; accepted 5 December 2014; advance access publication 14 January 2015.

Indices of abundance are the bedrock for stock assessments or empirical management procedures used to manage fishery catches for fish populations worldwide, and are generally obtained by processing catch-rate data. Recent research suggests that geostatistical models can explain a substantial portion of variability in catch rates via the location of samples (i.e. whether located in high- or low-density habitats), and thus use available catch-rate data more efficiently than conventional "design-based" or stratified estimators. However, the generality of this conclusion is currently unknown because geostatistical models are computationally challenging to simulation-test and have not previously been evaluated using multiple species. We develop a new maximum likelihood estimator for geostatistical index standardization, which uses recent improvements in estimation for Gaussian random fields. We apply the model to data for 28 groundfish species off the U.S. West Coast and compare results to a previous "stratified" index standardization model, which accounts for spatial variation using post-stratification of available data. This demonstrates that the stratified model generates a relative index with 60% larger estimation intervals than the geostatistical model. We also apply both models to simulated data and demonstrate (i) that the geostatistical model has well-calibrated confidence intervals (they include the true value at approximately the nominal rate), (ii) that neither model on average under- or overestimates changes in abundance, and (iii) that the geostatistical model has on average 20% lower estimation errors than a stratified model. We therefore conclude that the geostatistical model uses survey data more efficiently than the stratified model, and therefore provides a more cost-efficient treatment for historical and ongoing fish sampling data.

Keywords: abundance index, delta-generalized linear mixed model, fishery-independent data, Gaussian random field, geostatistics, index standardization, management procedure, spatial statistics, stock assessment, template model builder.

Introduction

Fisheries management throughout the United States, Europe, and elsewhere is often informed by estimates of fish population abundance derived from sampling operations with predetermined sampling designs. Data derived from fishery-independent surveys are generally processed to generate an index that is intended to be proportional to population abundance (Maunder and Punt, 2004). Abundance indices are generally considered to be the most important source of information regarding fishing impacts on marine populations (Francis, 2011), and are used in data-rich and data-limited stock assessments to inform changes in management decisions, e.g. allowable catches (Methot *et al.*, 2014).

Fisheries management agencies worldwide use many different methods to estimate abundance indices and there can be wide variation in methodology even within a single fisheries management agency. Using NOAA Fisheries in the United States as an example,

the Alaska and Northeast Fisheries Science Centers of the U.S. National Marine Fisheries Service primarily use design-based estimators, which generally calculate average catch rates within each predetermined sampling stratum in the sampling design, then generate an area-weighted sum of abundance in each stratum (Smith, 1990). Other scientific groups, including the Northwest Fisheries Science Center, use a delta-generalized linear mixed model ("delta-GLMM") to preprocess survey data while controlling for confounding effects (i.e. differences in fishing power among contracted sampling vessels; Helser *et al.*, 2004; Thorson and Ward, 2014). This delta-GLMM approach separately analyses positive catches ("encounters"), and the catch rates of positive catches ("positive catch rates"), then combines these two components in a final estimate of population abundance (Lo *et al.*, 1992; Stefansson, 1996). Scientific groups may also analyse survey data while controlling for auxiliary information, e.g. the placement of longline hooks (Bigelow and Maunder, 2007) or associations of co-occurring species (Stephens and MacCall, 2004).

Recent research suggests that geostatistical approaches to processing survey data may yield more precise and accurate indices of abundance than either design-based or conventional delta-GLMMs (Shelton *et al.*, 2014). Geostatistical models for survey data proceed by specifying that population densities at nearby sites are more similar than densities at geographically remote sites. This statistical assumption allows a geostatistical model to estimate a smoothed surface representing spatial variation in density. The model can then incorporate measured variables (i.e. bottom substrate type) to assess what proportion of spatial variation in density can be explained by habitat features. In contrast, conventional design-based models assume that average density is fixed within a given stratum, and variation among samples in a stratum contributes to increased estimation variance for average density and therefore population abundance (Petitgas, 2001). Shelton *et al.* (2014) showed that a large portion of variation in survey catch rates can be explained by whether the randomized location for a given sample happened to fall in good or poor habitat for darkblotched rockfish (*Sebastes crameri*), i.e. by within-stratum variation. By explaining a greater portion of variation in survey catch rates, the geostatistical model for darkblotched then had greater precision, and avoided spikes in abundance that were implausible for this long-lived species (Gertseva and Thorson, 2013).

However, it remains unknown whether geostatistical models are likely in general to have improved performance relative to other model types when estimating abundance indices. For example, Yu *et al.* (2013) compare performance of geostatistical and generalized linear models when analysing simulated data, and find that the geostatistical model may have decreased performance given the sampling design available for Lake Erie. We therefore develop a fast and generic tool for implementing the geostatistical approach to index standardization. This approach is then applied to data for 28 groundfish off the U.S. West Coast, and results are compared with an existing delta-GLMM model that uses multiple spatial strata to account for spatial variation in population densities. We also conduct a simulation exercise to explore the statistical properties of the geostatistical approach, including whether its confidence intervals are well calibrated and whether the geostatistical model under- or overestimates interannual variation in population abundance.

Methods

Model development

The goal of applying a geostatistical model to data from fisheries research surveys is to explain the catches of each species recorded in the survey, and hence to infer population density throughout the domain of the survey design. We use a delta-generalized linear mixed modelling framework (Lo *et al.*, 1992; Stefansson, 1996; Martin *et al.*, 2005), which separately models the probability of having non-zero catches ("encounters") and catch rates for each encounter ("positive catch rates"). The probability that a sample encounters the target species (i.e. that catch $C > 0$) is approximated via a first model component:

$$\Pr[C > 0] = p, \tag{1}$$

where p is the probability of encounter (see Tables 1 and 2 for list of notation used). Positive catches are then approximated via a second model component:

$$\Pr[C = c | C > 0] = \text{Gamma}(c, \sigma^{-2}, \lambda\sigma^2), \tag{2}$$

where $\text{Gamma}(c, x, y)$ is the value of a probability density function evaluated at value c given a gamma distribution with shape x and scale y, λ is the expected catch given that the species is encountered, and σ is the coefficient of variation of measurement errors for positive catch rates. We use the Gamma distribution here because of its flexibility, although we note that many other distributions can be explored for positive catch rates (Thorson and Ward, 2013).

Each of these two components are estimated here using Gaussian Markov random fields. A random field defines the probability of a given function (e.g. densities as a function of latitude and longitude), and is analogous to a conventional random variable, which defines the probability of a given variable (Rasmussen and Williams, 2006). Specifically, a random field defines the expected value, variance and covariance of a multivariate realization from a stochastic process. In this case, the stochastic process represents the aggregate impact of environmental and biological factors that are not directly observed but still contribute to the distribution and density of the target species (Shelton *et al.*, 2014; Thorson *et al.*, 2015). Because the delta-GLMM framework involves two model components (probability of encounters and positive catch rates), we estimate unique random fields for each. For a Gaussian random field $\mathbf{E}$, the value of the random field at a single fixed location $s = \langle x, y \rangle$ (where x and y are the easting and northing for that location) follows a normal distribution, and the value of the random field at several (but a finite number of) locations $\mathbf{s} = \langle \mathbf{x}, \mathbf{y} \rangle$ (where the locations $\mathbf{s}$ are fixed and hence have no information about the value of the field) is a multivariate normal distribution:

$$\mathbf{E}[\mathbf{s}] \sim MN(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3}$$

where MN is a multivariate normal distribution, μ_i is the expected value at the ith location (in the following μ_i is fixed at 0), and $\boldsymbol{\Sigma}$ is the covariance of the random field E_i at each location s_i. We specify that this covariance follows a Matérn distribution (with smoothness $\nu = 1$). We also specify that the covariance between locations s and s' is stationary, but include the potentially impact of geometric anisotropy, such that any linear transformation in units measuring location (i.e. a rotation or rescaling of one spatial

Table 1. List of all parameters for either the geostatistical model (Model "G"), the Bayesian stratified model (Model "S"), or both.

Parameter name	Symbol	Type	Model
Encounter probability	p_i	Derived	G, S
Positive catch rates	λ_i	Derived	G, S
Anisotropy matrix	$\mathbf{H}$	Derived	G
Northings anisotropy	h_1	Fixed	G
Anisotropic correlation	h_2	Fixed	G
Average reference density (encounters/positive catch rates)	$d_t^{(p)}/d_t^{(\lambda)}$	Fixed	G, S
Covariate effects (encounters/positive catch rates)	$\beta_t^{(p)}/\beta_t^{(\lambda)}$	Fixed	G, S
Vessel effects (encounters)	$r_t^{(p)}/r_t^{(\lambda)}$	Random	G, S
Spatial residuals (encounters/positive catch rate)	$\omega_i^{(p)}/\omega_i^{(\lambda)}$	Random	G
Spatio-temporal residuals (encounters/positive catch rate)	$\varepsilon_{i,t}^{(p)}/\varepsilon_{i,t}^{(\lambda)}$	Random	G
Variation among strata and years (encounters/positive catch rate)	$\gamma_{s,t}^{(p)}/\gamma_{s,t}^{(\lambda)}$	Random	S
Average variation among strata (encounters/positive catch rate)	$\mu_s^{(p)}/\mu_s^{(\lambda)}$	Random	S
Variance of vessel effects (encounters/positive catch rate)	$\sigma_r^{2(p)}/\sigma_r^{2(\lambda)}$	Fixed	G, S
Variation of spatial residuals (encounters/positive catch rate)	$\tau_\omega^{(p)}/\tau_\omega^{(\lambda)}$	Fixed	G
Variation of spatio-temporal residuals (encounters/positive catch rate)	$\tau_\varepsilon^{(p)}/\tau_\varepsilon^{(\lambda)}$	Fixed	G
Covariance of spatial residuals (encounters/positive catch rate)	$\boldsymbol{\Sigma}_\omega^{(p)}/\boldsymbol{\Sigma}_\omega^{(\lambda)}$	Derived	G
Covariance of spatiatemporal residuals (encounters/positive catch rate)	$\boldsymbol{\Sigma}_\varepsilon^{(p)}/\boldsymbol{\Sigma}_\varepsilon^{(\lambda)}$	Derived	G
Range of spatial and spatio-temporal residuals (encounters/positive catch rate)	$\kappa^{(p)}/\kappa^{(\lambda)}$	Fixed	G
Variation among strata and years (encounters)	$\sigma_\gamma^{(p)}/\sigma_\gamma^{(\lambda)}$	Fixed	S
Coefficient of variation of measurement errors (positive catch rates)	Σ	Fixed	G, S

The type of each parameter is listed as estimated ("fixed", "random"), or calculated from estimated parameters ("derived").

Table 2. List of indices used in model descriptions, and data used during parameter estimation.

Name	Symbol	Type	Model
Sample	i	Index	G, S
Knot	j	Index	G
Covariate	k	Index	G, S
Stock	l	Index	G, S
Year	t	Index	G, S
Stratum	s	Index	S
Area swept for sample i	w_i	Data	G, S
Catch (in kilograms) for sample i	c_i	Data	G, S
Survey vessel for sample i	$V(i)$	Data	G, S
Nearest knot for sample i	$J(i)$	Data	G
Year for sample i	$T(i)$	Data	G, S
Stratum for sample i	$S(i)$	Data	S
Area associated with knot j	a_j	Data	G
Area associated with stratum s	a_s	Data	S
Covariate k at knot j	$x_{j,k}$	Data	G, S

axis but not the other) will have no impact upon model results:

$$\Sigma(s, s') = \sigma_E^2 \cdot \text{Matérn}(\|\mathbf{H}(s - s')\|), \quad (4)$$

where $\mathbf{H}$ is a linear transformation representing geometric anisotropy, $(s - s') = \langle x - x',\ y - y' \rangle$ is the difference in eastings and northings between locations s and s', and $||\mathbf{H}(s - s')||$ is the distance between locations after accounting for geometric anisotropy (see Cressie and Wikle, 2011, Eq. 4.9 for details). The matrix $\mathbf{H}$ is defined to preserve volume (i.e. has a determinant of one), and is calculated from two estimated parameters (see Supplementary Appendix A).

For computational reasons, we use a "predictive" framework for spatio-temporal models, where the value of a random field $\mathbf{E}$ defined over a domain Ω is approximated as being piecewise constant. To accomplish this, the user pre-specifies the desired number of "knots" n_j that are used to approximate random field $\mathbf{E}$ over domain Ω, such that the model tracks the value of $\mathbf{E}$ at each knot. The value of $\mathbf{E}$ at any location s is equal to its value at the nearest knot (hence the piecewise-constant approximation). Therefore, the value of $\mathbf{E}$ at location s_i (the location for the ith sample) is determined from the value $E_{J(i)}$ at the knot $J(i)$ that is nearest to s_i. The location of all n_j knots is determined by applying a cluster algorithm (in this case, a k-means algorithm) to the locations of available sampling data. This clustering algorithm results in a distribution of knots with density that is proportional to the sampling intensity of the survey design. The area a_j associated with each knot j is then calculated using the Voronoi tool in the *PBSmapping* package in R (Schnute *et al.*, 2013). After being calculated, the number and location of all knots is held fixed during parameter estimation. The prespecified number of knots controls the accuracy of this piecewise-constant approximation, and can be used to achieve a balance of accuracy and computational speed. This approximation also simplifies calculating the integral across the random field, as is necessary when calculating the index of abundance (as discussed in detail later).

We approximate encounter probability p and positive catch rates λ using a link function and a combination of linear predictors (including random fields). Linear predictors represent average density d_t in each year t, the relative fishing efficiency r_v for the vth survey vessel, the association β_x of any measured environmental variable x with encounter probabilities or positive catch rates, spatially correlated variability ω_j in encounter probabilities or positive catch rates at a knot j that is persistent among years, and spatially correlated variability $\varepsilon_{j,t}$ at a given knot j in year t. We specifically use the following model for encounter probability p_i for sample i at location s_i:

$$p_i = \text{logit}^{-1}\left(d_{T(i)}^{(p)} + \sum_{k=1}^{n_x} \beta_k^{(p)} x_{J(i),k} + r_{V(i)}^{(p)} + \omega_{J(i)}^{(p)} + \varepsilon_{J(i),T(i)}^{(p)}\right), \quad (5)$$

where $T(i)$ is the year for sample i, $V(i)$ is the vessel for sample i, $J(i)$ is the nearest knot to sample i (where we use capital letters

to represent indicator variables T, V, J, and S, and S is defined later; see Table 2), and n_x is the number of measured covariates that are included in the model. Similarly, we approximate positive catch rates λ_i for sample i:

$$\lambda_i = w_i \cdot \exp\left(d^{(\lambda)}_{T(i)} + \sum_{k=1}^{n_x} \beta^{(\lambda)}_k x_{J(i),k} + r^{(\lambda)}_{V(i)} + \omega^{(\lambda)}_{J(i)} + \varepsilon^{(\lambda)}_{J(i),T(i)} \right), \quad (6)$$

where w_i is the area swept for sample i. For both model components, vessel effects $\boldsymbol{r}$, spatial residuals $\boldsymbol{\omega}$ and spatio-temporal residuals $\boldsymbol{\varepsilon}$ are all random. For the encounter probability component,

$$\begin{aligned} r^{(p)}_v &\sim N(0, \sigma^{2(p)}_r) \\ \boldsymbol{\omega}^{(p)} &\sim MN(0, \boldsymbol{\Sigma}^{(p)}_\omega) \\ \boldsymbol{\varepsilon}^{(p)}_t &\sim MN(0, \boldsymbol{\Sigma}^{(p)}_\varepsilon) \end{aligned} \quad (7)$$

and random effects for positive catch rates are defined in the same way.

Our model therefore includes multiple sources of stochastic variation and can be conceptualized hierarchically (Searle *et al.*, 1992; Cressie *et al.*, 2009; Thorson and Minto, 2014). Specifically, this model is based on the interpretation that population density across sampling domain $\boldsymbol{\Omega}$ in a given year is a realizations of a random function. This random functions arises from the combination of measured covariates $\mathbf{x}$ and random fields $\boldsymbol{\omega}$ and $\boldsymbol{\varepsilon}$, and assumptions about these random fields (i.e. that they stationary but have geometric anisotropy) result in the properties of the random function used to describe spatial variation in population density. The stochastic process of sampling local densities using bottom trawl gear then introduces another source of variability, which follows a Bernoulli or gamma distribution (Eqs. (1) and (2)), and vessels have random variation in catch efficiency, thus representing a third source of variability. The treatment of space via random functions allows for inference similar to that for "intrinsic" geostatistics, such that the geostatistical model (Eqs. (1)–(7)) can also be applied to non-randomized sampling designs (Petitgas, 2001), subject to the assumption the process of selecting sampling locations is independent of the process generating difference in population density (Diggle *et al.*, 2010).

We estimate fixed effects via maximum marginal likelihood while integrating across all random effects (see Table 1 for full list of parameters). The marginal likelihood is approximated using the Laplace approximation (Skaug and Fournier, 2006; Thorson *et al.*, 2015) as implemented using template model builder (Kristensen *et al.*, 2014). The conditional probability of random effects was approximated using the stochastic partial differential equation approach for stationary, geometric anisotropy outlined in Supplementary Appendix A (Lindgren *et al.*, 2011). The marginal likelihood was then maximized using conventional non-linear optimization in the R statistical platform (R Core Development Team, 2013). Further details regarding this computational approach for spatio-temporal models can be found in Thorson *et al.*, (2015).

In our West Coast application, we proceed by obtaining a grid that encompasses the entire spatial domain for the available survey data. Each grid cell is 2 km × 2 km, which results in ~40 000 grid cells. For each grid cell, we know the eastings and northings at the centroid, and the average value of every covariate (x_k in $\mathbf{X}$). The value of the random field in each cell is assumed to be equal to its value at the nearest knot (i.e. following the piecewise-constant approximation), so the area a_j associated with the jth knot is calculated as the number of grid cells associated with it times their areas. The total abundance across the entire population domain can then be calculated as follows:

$$\begin{aligned} \hat{b}_t = \sum_{j=1}^{n_j} a_j \mathrm{logit}^{-1}\left(\hat{d}^{(p)}_t + \sum_{k=1}^{n_x} \hat{\beta}^{(p)}_k x_{j,k} + \hat{\omega}^{(p)}_j + \hat{\varepsilon}^{(p)}_{j,t} \right) \exp \\ \left(\hat{d}^{(\lambda)}_t + \sum_{k=1}^{n_x} \hat{\beta}^{(\lambda)}_k x_{j,k} + \hat{\omega}^{(\lambda)}_j + \hat{\varepsilon}^{(\lambda)}_{j,t} \right), \end{aligned} \quad (8)$$

where $\hat{b}_t$ is the estimate of total abundance in year t, $\hat{d}^{(p)}_t$, $\hat{\beta}^{(p)}_k$, $\hat{d}^{(\lambda)}_t$, and $\hat{\beta}^{(\lambda)}_k$ are fixed effects estimated via maximum likelihood, and $\hat{\omega}^{(p)}_t$, $\hat{\varepsilon}^{(p)}_{j,t}$, $\hat{\omega}^{(\lambda)}_t$, and $\hat{\varepsilon}^{(\lambda)}_{j,t}$ are random effects set to the value that maximizes the joint likelihood conditional on the estimated value of fixed effects. Total abundance is calculated by summing across predicted density for all knots, where each is weighted by its area a_j. This "area-weighting" is standard when calculating design-based indices (Cochran, 1977), and ensures that variation in sampling intensity does not influence the weighting assigned to different segments of the population (which could otherwise result in biased estimates of population abundance). Exploratory analysis showed that confidence intervals are more symmetric (and closer to quadratic; Bolker *et al.* (2013)) in log-space, so we report standard errors for $\log(b_t)$. Standard error estimates $\widehat{SE}[\log(\hat{b}_t)]$ are computed via a first-order Taylor series expansion (sometimes called the generalized delta-method; see Fournier *et al.*, 2012 for details) by template model builder (Kristensen *et al.*, 2014; available at https://github.com/kaskr/adcomp), while $\hat{b}_t$ is calculated via a plug-in estimate. We note that future analysis could explore alternative methods for summarizing uncertainty, e.g. likelihood profile or bootstrap methods (Magnusson *et al.*, 2013), but we do not do so here. We authors are conducting ongoing research regarding improved estimators for derived quantities such as $\hat{b}_t$ in maximum likelihood mixed-effects models (e.g. Tierney *et al.* (1989)), in place of the plug-in estimator used in this paper, but the topic is not explored further here. Code for adapting this geostatistical delta-GLMM is provided on GitHub (https://github.com/nwfsc-assess/geostatistical_delta-GLMM).

Comparison with stratified delta-GLMM

We compare the geostatistical delta-GLMM with results from a Bayesian delta-GLMM that includes spatial stratification to account for spatial variation in densities, and which is conventionally used for these data (Thorson and Ward, 2013, 2014). This stratified delta-GLMM accounts for spatial variation in encounter probabilities and positive catch rates by dividing the spatial domain into 15 spatial strata, using three divisions by depth (55–183, 184–549, and 550–1280 m) and five latitudinal divisions (32–34.5 N, 34.5–40.5 N, 40.5–43 N, 43–47.5 N, and 47.5–50 N), and eliminating all data for any stratum that has three or fewer encounters across all 10 years of data. These spatial strata include strata that are included in the survey design (with small variation in sampling intensity among strata), as well as additional post-stratification as is conventionally done for these data. Within each stratum, encounter probabilities and positive catch rates are assumed to be uniform when using the stratified delta-GLMM.

The stratified delta-GLMM estimates encounter probability p_i as

$$p_i = \text{logit}^{-1}\left(d^{(p)}_{T(i)} + \sum_{k=1}^{n_x} \beta^{(p)}_k x_{i,k} + \mu^{(p)}_{S(i)} + r^{(p)}_{V(i)} + \gamma^{(p)}_{S(i),T(i)}\right), \quad (9)$$

where $\mu^{(p)}_{S(i)}$ is a fixed effect representing average encounter probabilities for the stratum $S(i)$ containing sample i (for identifiability $\mu^{(p)}_1 = 0$), and $\gamma^{(p)}_{S(i),T(i)}$ is a random effect representing variation in encounter probability for stratum $S(i)$ and year $T(i)$ (Thorson and Ward, 2013):

$$\gamma^{(p)}_{s,t} \sim N(0, \sigma^{2(p)}_\gamma), \quad (10)$$

where $\sigma^{(p)}_\gamma$ is the estimated standard deviation of variation among strata and years. Similarly, we approximate positive catch rates λ_i for sample i:

$$\lambda_i = w_i \cdot \exp\left(d^{(\lambda)}_{T(i)} + \mu^{(\lambda)}_{S(i)} + \sum_{k=1}^{n_x} \beta^{(\lambda)}_k x_{S(i),k} + r^{(\lambda)}_{V(i)} + \gamma^{(\lambda)}_{S(i),T(i)}\right), \quad (11)$$

where $\mu^{(\lambda)}_{s(i)}$ is a fixed effect and $\gamma^{(\lambda)}_{s(i),t(i)}$ is random, $\gamma^{(\lambda)}_{s,t} \sim N(0, \sigma^{2(\lambda)}_\gamma)$, and $\sigma^{(\lambda)}_\gamma$ is an estimated parameter. We then calculate expected biomass b_t in year t as

$$b_t = \sum_{s=1}^{n_s} a_s \text{logit}^{-1}(d^{(p)}_t + \mu^{(p)}_s + \gamma^{(p)}_{s,t}) \exp(d^{(\lambda)}_t + \mu^{(\lambda)}_s + \gamma^{(\lambda)}_{s,t}), \quad (12)$$

where n_s is the number of strata (15 in the following case-study example). This equation again implies that strata are weighted by their area a_s, analogous to the weighting of knots by area a_j associated with each knot j in the geostatistical model. Equation (12) is used to calculate expected biomass for each sample from the posterior distribution of the model, and we then summarize biomass using the posterior median and standard deviation of $\log(b_t)$ (see Thorson and Ward, 2014 for more details).

In summary, the stratified model has coefficients for year and vessel that are similar to those in the geostatistical model, and differs primarily by replacing random fields for spatial variation (ω) and spatio-temporal variation (ε) with coefficients for strata (μ) and strata-year (γ) effects. Given that the stratified model is Bayesian, it requires specifying priors for all parameters. We specifically use a weakly informative prior on the coefficient of variation of measurement errors for positive catch rates (σ), $\Pr(1/\sigma^2) = \text{Gamma}(0.001,0.001)$, a uniform prior on the standard deviation of vessel and strata-year effects, $\Pr(X) = 1/100$ if $0 < X < 100$ and zero otherwise where X is σ_r and σ_γ, and bounded uniform priors on all other parameters $\Pr(X) = 1/40$ if $-20 < X < 20$ and zero otherwise where X represents d_t, μ_s, and β_k. The Bayesian model is then fitted using Just Another Gibbs Sampler (JAGS; Plummer (2003)) called from the R statistical platform. Code can be found at the Northwest Fisheries Science Center GitHub repository (*nwfscDeltaGLM*; https://github.com/nwfsc-assess/nwfscDeltaGLM), and further details can be found in Thorson and Ward (2014).

Case-study application

We apply the geostatistical and stratified delta-GLMMs to data derived from a multispecies bottom trawl survey for fish off the U.S. West Coast (Oregon, Washington, and California), operated with consistent sampling protocol by the Northwest Fisheries Science Center from 2003 to 2012 (Bradburn *et al.*, 2011). This design has variable but predetermined sampling intensity in different strata, and this will not affect the assumed models for the data given that the sampling intensity is independent of underlying variation in density (Diggle and Ribeiro, 2007, section 1.2.3). The design involves contracting with commercial fishers for labour and fishing vessels, and hence requires accounting for potential variation in fishing power between vessels (Helser *et al.*, 2004; Thorson and Ward, 2014). We specifically apply the model to data for 28 groundfish species, selected to represent a variety of life history types and including all previously assessed species (Thorson and Ward, 2013). Though covariates can be included in either the geostatistical model or stratified delta-GLMM (Thorson and Ward, 2013; Shelton *et al.*, 2014), we have omitted covariates from both models here (i.e. $\beta^{(p)} = \beta^{(\lambda)} = 0$) to isolate the comparison of models with and without spatial random fields. We use 1000 knots for all case-study runs of the geostatistical model (i.e. $n_j = 1000$). In contrast, the default stratification of the stratified model involves estimating differences in density among 15 strata. Spatial strata for the stratified model vary in area from 1.5 to 28.5 thousand km^2, such that the number of knots located within each spatial stratum varies from 18 to 149. Having multiple knots per stratum allows the geostatistical model to approximate spatial variation within each stratum, thus reducing residual (unexplained) variation in survey data (Shelton *et al.*, 2014). Specifying 1000 knots in the geostatistical model results in >24 000 estimated coefficients per species, and geostatistical models for all 28 species can run overnight on a single core of a laptop. Exploratory analysis illustrates that varying the number of knots has little impact on reported results.

We then compare the coefficient of variation for the geostatistical model, $CV^{(\text{geo})}_{l,t}$ for stock l in year t:

$$CV^{(\text{geo})}_{l,t} = \sqrt{\exp(SE(\log(\hat{b}^{(\text{geo})}_{l,t}))^2) - 1}, \quad (13)$$

where $\log(\hat{b}_{l,t})$ is the estimated log-biomass for stock l in year t, $SE(\log(\hat{b}_{l,t}))$ is the estimated standard error for log-biomass for that stock and year, and the coefficient of variation for the stratified model $\text{CV}^{(\text{strat})}_{l,t}$ is defined in the same way. Finally, we calculate the average ratio for the two models, $(1/28)(1/10)\sum_{l=1}^{28}\sum_{t=2003}^{2012} CV^{(\text{strat})}_{l,t}/CV^{(\text{geo})}_{l,t}$, and use this to evaluate whether the geostatistical or stratified model provides a more precise estimate of abundance.

Evaluation using simulated data

In addition to the case-study application, we also conduct a simulation experiment. In each replicate for this experiment, we simulate data similar to that available off the U.S. West Coast, and then fit both the stratified and geostatistical delta-GLMMs to the simulated data. We then extract the estimate of total abundance and its associated standard error from each model, and compare these with the true, simulated abundance for that replicate. We use 100 replicates for this experiment, and ensure that this sample size is sufficiently large that results are qualitatively similar when replicating the experiment.

Each simulation replicate proceeds by randomly selecting 600 of the ~40 000 cells that cover the entire sampling domain in each year

t, then simulating encounters and positive catch rates:

$$P_i \sim Bernoulli(p_i),$$
$$C_i \sim \begin{cases} \text{Gamma}(\sigma^{-2}, \lambda_i\sigma^2) & \text{if } P_i = 1 \\ 0 & \text{if } P_i = 0 \end{cases}, \quad (14)$$

where p_i and λ_i are generated using Eqs. (5)–(6) above. In the simulation, we use a covariate matrix $\mathbf{X} = \langle \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{x}^{(3)} \rangle$ where $\boldsymbol{x}^{(1)}$ is depth, $\boldsymbol{x}^{(2)}$ is depth squared, and $\boldsymbol{x}^{(3)}$ is the square-root of distance from nearest rocky habitat (Shelton *et al.*, 2014) and $\boldsymbol{\beta}^{(p)} = \boldsymbol{\beta}^{(\boldsymbol{\lambda})} = \langle 1, -1, 1 \rangle$ (i.e. depth has a quadratic effect on both encounter probabilities and positive catch rates, and both decline as a function of distance from nearest rocky habitat). The random fields used to simulate these data ($\boldsymbol{\omega}^{(p)}$, $\boldsymbol{\varepsilon}^{(p)}$, $\boldsymbol{\omega}^{(\lambda)}$, and $\boldsymbol{\varepsilon}^{(\lambda)}$) are specified to have a marginal standard deviation of one, and have a stationary, isotropic correlation matrix ($\mathbf{H} = \mathbf{I}$) where the range (distance at which correlations decline to 0.1) is 1000 km. for encounter probabilities and 500 km for positive catch rates. Simulated random fields had a Gaussian distance function, and were generated using the *RandomFields* package (Schlather, 2009) in R.

We use these simulated data to compare the performance of stratified and geostatistical delta-GLMMs in three ways:

(i) First, we address whether either model is losing information regarding interannual variation in abundance. Specifically, we run the following linear model for each of the two models:

$$\log(\hat{b}_{r,t}) = \alpha_r + \delta \log(b_{r,t}) + \varepsilon_{r,t},$$
$$\varepsilon_{r,t} \sim N(0, \sigma_\varepsilon^2), \quad (15)$$

where $b_{r,t}$ is true abundance for replicate r in year t, $\hat{b}_{r,t}$ is estimated abundance for that replicate and year, α_r is a nuissance parameter for each replicate (reflecting random differences in scale between true and estimated indices), $\varepsilon_{r,t}$ is "estimation error" in the estimated index, and δ reflects whether the estimated index reflects changes in true abundance accurately ($\delta = 1$), underestimates changes in abundance ($\delta < 1$), or overestimates changes in abundance ($\delta > 1$; see Wilberg *et al.* (2010)). In this context, then, δ can be interpreted as a measure of model bias, where a model where $\delta < 1$ will have result in an estimate of b_t that is biased towards the estimate of b_{t-1}.

(ii) We also compare the magnitude of errors for the geostatistical and stratified models when estimating relative abundance:

$$\text{Error}_{r,t} = \hat{I}_{r,t} - I_{r,t}, \quad (16)$$

where $\hat{I}_t = \hat{b}_t / \exp((1/n_t)\sum_{t=i}^{n_t} \log(\hat{b}_t))$ and $\hat{b}_t$ is estimated abundance, $I_t = b_t / \exp((1/n_t)\sum_{t=i}^{n_t} \log(b_t))$ and b_t is true abundance. We then calculate root-mean-squared error separately for results from the geostatistical and stratified models:

$$RMSE = \sqrt{\frac{1}{100}\frac{1}{10}\sum_{r=1}^{100}\sum_{t=1}^{10} \text{Error}_{r,t}^2}, \quad (17)$$

which is a measure of estimation error for each model. Specifically, calculating an index of abundance (Eq. (8)) involves predicting biomass across the survey domain $\boldsymbol{\Omega}$ (or, in practice, predicting biomass at each knot j). A model with too many parameters will result in imprecise predictions of biomass, i.e. decreased precision for estimated indices of abundance, and will therefore have a greater *RMSE* than a model that is more parsimonious (Burnham and Anderson, 2002).

(iii) Finally, we explore whether the geostatistical model is generating estimated confidence intervals that accurately represent estimation uncertainty. To do so, we calculate the quantile distribution for each year and replicate:

$$Q_{r,t} = \int_{-\infty}^{I_{r,t}} N(\log(\hat{I}_{r,t}), \widehat{SE}(\log(\hat{I}_{r,t}))) \, d\hat{I}_{r,t} \quad (18)$$

where $I_{r,t} = b_{r,t} / \exp\left((1/n_t)\sum_{t=i}^{n_t} \log(b_{r,t})\right)$ is the true relative abundance, $\hat{I}_{r,t} = \hat{b}_{r,t} / \exp((1/n_t)\sum_{t=i}^{n_t} \log(\hat{b}_{r,t}))$ is the estimated relative abundance, and $\widehat{SE}(\log(\hat{I}_{r,t}))$ is the estimated standard error for $\log(\hat{I}_t)$. A well-calibrated confidence interval will have quantiles that approximately follow a uniform distribution from 0.0 to 1.0, while a poorly calibrated confidence interval may be either overdispersed (too many values close to 0.5) or underdispersed (too many values close to 0.0 or 1.0) relative to a standard uniform distribution.

Results

We first show estimates of population density (log-kg km^{-2}) for an example species, arrowtooth flounder (*Atheresthes stomias*; Figure 1). Log-densities range widely from −5 to 10 across the spatial domain of the survey data, and are highest near Washington. However, densities appear to increase south of the Sacramento Delta after 2009. Changes in the proportion of the population in different spatial areas are estimated via the annual random field (i.e. $\varepsilon^{(p)}$ and $\varepsilon^{(\lambda)}$), and both processes are responsible for this variation for arrowtooth flounder ($\sigma_\varepsilon^{(p)} = 0.82$; $\sigma_\varepsilon^{(\lambda)} = 0.59$; Table 3). In general, however, spatial variation that is constant over time (i.e. $\omega^{(p)}$ and $\omega^{(\lambda)}$) has a greater magnitude than spatial variation that changes annually for these 28 species (median across species: $\sigma_\omega^{(p)} = 4.23$; $\sigma_\omega^{(\lambda)} = 1.52$; $\sigma_\varepsilon^{(p)} = 0.45$; $\sigma_\varepsilon^{(\lambda)} = 0.74$; Table 3), and spatial variation is much larger than differences in sampling efficiency among contracted fishing vessels ($\sigma_r^{(p)} = 0.08$; $\sigma_r^{(\lambda)} = 0.11$).

We also display the estimated anisotropy for each species for the encounter probability and positive catch-rate model components (Figure 2). For arrowtooth flounder, this shows that spatial residuals in positive catch rates (ellipse with solid line) are correlated over a longer distance than are residual encounter probabilities (the ellipse for positive catch rates is larger than the ellipse for encounter probabilities), and the correlation is oriented

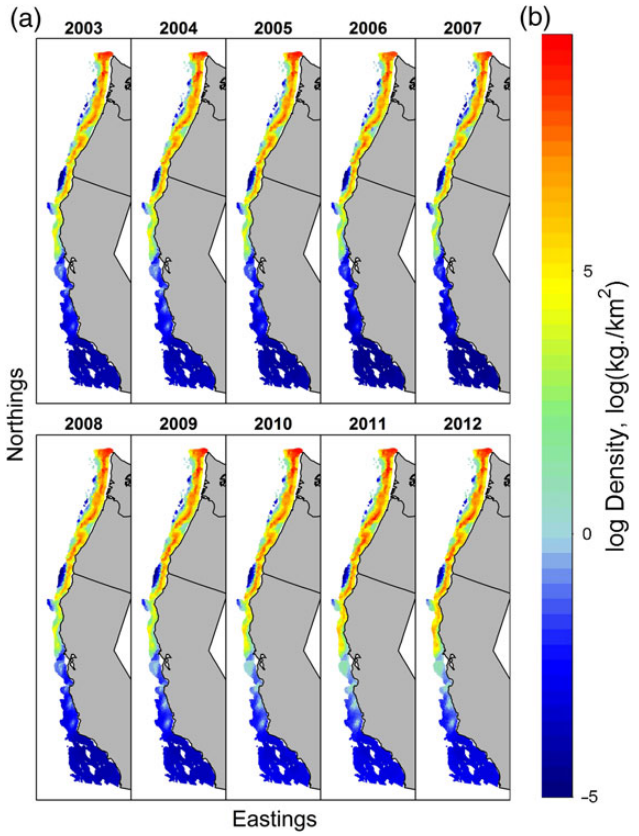

Figure 1. Density for arrowtooth flounder 2003–2012, estimated by the geostatistical delta-generalized linear mixed model (note that the white space in southern California represents the cowcod conservation area, which prohibits trawl gears including the survey design and hence is excluded when estimating spatial densities and abundance indices).

North—South (the ellipse for arrowtooth is stretched along the North–South axis), i.e. along the shoreline along the Oregon and Washington coasts. Many other species have a correlation in spatial residuals that declines slower North–South than East–West, and the distance at which correlations drop to ~0.1 varies greatly among species from <10 km (yelloweye) to >600 km (cowcod and yellowtail).

We next compare the estimated indices of relative abundance for the geostatistical and the stratified model (Figure 3). In general, the geostatistical and stratified indices show similar trends, e.g. decreases for Pacific hake and English sole, and increases for halfbanded and petrale sole. However, in many cases the 95% interval for the stratified index is noticeably wider than that for the geostatistical index, e.g. canary, arrowtooth, greenstriped, rosethorn, and yellowtail. In contrast, there are relatively few instances where the interval for the geostatistical model is wider than the stratified model (although see 2007 of rougheye, 2011 of Pacific Ocean perch, etc.). This observation is borne out when comparing the coefficient of variation for the geostatistical and stratified models (Figure 4). On average, the stratified model has a coefficient of variation (CV) that is 60% larger than the CV for the geostatistical model.

Finally, we summarize results for the 100 replicates of the simulation experiment (Figures 5 and 6). This demonstrates that neither the stratified nor the geostatistical model substantially over- or underestimate changes in true abundance on average (Eq. (15); $\delta^{(\text{strat})} = 1.00$; $\delta^{(\text{geo})} = 1.01$). However, the stratified model has nearly 20% greater error ($RMSE = 0.127$) than the geostatistical model ($RMSE = 0.109$). The quantile distribution for the geostistical model is slightly overdispersed, showing that the estimated standard errors might be slightly too small for these simulated data (Figure 6). In contrast, the quantile disribution for the stratified model is slightly underdispersed, implying that the stratified model might have credible intervals that are slightly too wide.

Discussion

We have developed a geostatistical model that estimates spatial variation in encounter probabilities and positive catch rates, and which can be used to estimate an index of population abundance. We envision that, if the geostatistical index were included in a stock assessment model, the geostatistical index of abundance $\log(b_t)$ would be treated as proportional to the logarithm of available biomass B_t, with residual variance equal to the square of the estimated standard error plus a variance-inflation term σ_b^2:

$$\log(B_t) = \log(q) + \log(\hat{b}_t) + e_t, \tag{19}$$

$$e_t \sim N(0, \widehat{SE}(\log(\hat{b}_t))^2 + \sigma_b^2),$$

where the variance-inflation accounts for the magnitude of random variation in the catchability coefficient q among years (Wilberg *et al.*, 2010). This treatment conforms to the usage of existing abundance index estimates for the U.S. West Coast (Methot and Wetzel, 2013; Thorson and Ward, 2013). In this case, the standard error of the log-index, $\widehat{SE}(\log(\hat{b}_t))$, represents an upper-bound on the weighting that is given to the geostatistical index relative to other data that are included in the assessment model (see Francis, 2011 for a full discussion).

In this geostatistical model, spatial variation is decomposed into a component that is constant across time, and another component that varies among years, such that the model can account for changes in spatial distribution over time (e.g. for arrowtooth flounder). Spatial distributions for assessed species will change whenever species exhibit density-dependent habitat selection (MacCall, 1990), and density-dependent habitat selection has previously been shown to cause bias in relative indices of abundance using conventional stratified estimators for survey data (Thorson *et al.*, 2013). Improving upon past developments in Shelton *et al.* (2014), we also account for geometric anisotropy (where densities are more correlated the direction parallel with shore than perpendicular to shore) and by including random variation caused by contracting multiple fishing vessels within and among years (Thorson and Ward, 2014).

Using this new statistical approach, we have shown that the stratified model has 60% larger credible intervals than the geostatistical delta-GLMM for 28 groundfish species. A simulation experiment demonstrates that the stratified model has overdispersed credible intervals (they cover the true value more than the nominal amount), while the geostatistical model has slightly underdispersed confidence intervals (they cover the true value less than the nominal amount). However, neither effect appears to have a large magnitude for the simulated data, so it seems unlikely to account for the 60% difference in interval width seen for West Coast groundfish. Additionally, the simulation experiment demonstrates that the stratified model has a 20% greater simulation error.

Table 3. Estimates of variance parameters for all 28 West Coast groundfish analysed in the case-study application (for computation of marginal standard deviation for random fields, see Supplementary Appendix A).

Species name		Random fields (marginal SD)				Vessel effects (SD)		
Common	Scientific	$\sigma_{\varepsilon}^{(p)}$	$\sigma_{\varepsilon}^{(\lambda)}$	$\sigma_{\omega}^{(p)}$	$\sigma_{\omega}^{(\lambda)}$	$\sigma_{r}^{(p)}$	$\sigma_{r}^{(\lambda)}$	Residual. error $\boldsymbol{\sigma}$
Median	–	0.451	0.742	4.233	1.523	0.080	0.111	1.062
Arrowtooth	*Atheresthes stomias*	0.815 (0.140)	0.594 (0.069)	8.658 (1.475)	2.263 (0.272)	0.193 (0.089)	0.161 (0.038)	0.966 (0.016)
Aurora	*Sebastes aurora*	0.452 (0.140)	0.259 (0.123)	4.211 (0.416)	1.343 (0.192)	0.105 (0.141)	0.069 (0.104)	1.050 (0.028)
Bocaccio	*S. paucispinis*	0.712 (0.158)	0.879 (0.154)	3.741 (0.473)	0.697 (0.197)	0.007 (0.177)	0.007 (0.099)	1.127 (0.054)
Canary	*S. pinniger*	0.001 (0.170)	1.125 (0.132)	3.486 (0.587)	1.203 (0.178)	0.082 (0.170)	0.135 (0.226)	0.988 (0.052)
Chilipepper	*S. goodie*	0.601 (0.143)	1.538 (0.130)	4.397 (0.468)	2.036 (0.238)	0.007 (0.055)	0.314 (0.127)	1.396 (0.041)
Cowcod	*S. veils*	0.001 (0.206)	0.003 (0.343)	3.729 (0.825)	2.197 (0.401)	0.225 (0.171)	0.007 (0.196)	1.150 (0.066)
Darkblotched	*S. crameri*	0.000 (0.172)	0.939 (0.081)	4.062 (0.572)	1.686 (0.168)	0.079 (0.130)	0.340 (0.074)	1.149 (0.031)
Dover sole	*Microstomus pacificus*	0.677 (0.120)	0.436 (0.047)	4.103 (0.438)	2.831 (0.226)	0.128 (0.088)	0.007 (0.061)	1.021 (0.010)
English sole	*Parophrys vetulus*	0.608 (0.114)	0.908 (0.062)	10.053 (1.521)	1.352 (0.137)	0.222 (0.071)	0.111 (0.039)	0.971 (0.017)
Greenspotted	*S. chlorostictus*	0.001 (0.175)	0.729 (0.285)	3.740 (0.456)	1.587 (0.225)	0.007 (0.088)	0.007 (0.578)	1.176 (0.061)
Greenstriped	*S. elongates*	0.518 (0.115)	0.839 (0.078)	4.939 (0.622)	1.824 (0.142)	0.165 (0.071)	0.111 (0.075)	1.193 (0.025)
Halfbanded	*S. semicintus*	0.001 (0.112)	1.556 (0.214)	6.819 (1.086)	2.339 (0.282)	0.007 (0.119)	0.179 (0.267)	1.475 (0.054)
Hake	*Merluccius productus*	0.001 (0.087)	0.002 (0.044)	16.511 (3.116)	2.605 (0.571)	0.007 (0.143)	0.032 (0.055)	0.829 (0.012)
Longspine thornyhead	*Sebastolobus altivelis*	0.879 (0.084)	1.048 (0.056)	4.338 (0.647)	0.991 (0.080)	0.331 (0.058)	0.128 (0.045)	1.232 (0.018)
Petrale	*Eopsetta jordani*	0.000 (0.162)	0.754 (0.152)	4.255 (0.747)	1.720 (0.226)	0.018 (0.914)	0.010 (0.659)	1.139 (0.057)
Pacific Ocean perch	*S. alutus*	0.553 (0.142)	0.632 (0.088)	12.701 (2.016)	2.425 (0.254)	0.166 (0.107)	0.166 (0.040)	1.069 (0.019)
Redbanded	*S. babcocki*	0.004 (0.129)	0.482 (0.056)	9.391 (1.483)	1.238 (0.136)	0.007 (0.081)	0.111 (0.032)	0.937 (0.014)
Rosethorn	*S. helvomaculatus*	0.000 (0.127)	0.729 (0.135)	3.376 (0.469)	0.943 (0.208)	0.181 (0.107)	0.007 (0.088)	1.033 (0.039)
Rougheye	*S. aleutianus*	0.000 (0.176)	0.003 (0.234)	2.742 (0.297)	1.460 (0.134)	0.022 (0.646)	0.249 (0.118)	1.173 (0.041)
Sablefish	*Anoplopoma fimbria*	0.002 (0.113)	0.001 (0.191)	3.230 (0.547)	2.255 (0.648)	0.007 (0.106)	0.007 (0.089)	1.019 (0.047)
Sanddab	*Citharichthys sordidus*	0.908 (0.102)	0.564 (0.043)	3.518 (0.360)	1.279 (0.094)	0.330 (0.059)	0.115 (0.026)	0.928 (0.012)
Sharpchin	*S. zacentrus*	0.300 (0.250)	1.877 (0.386)	2.953 (0.377)	2.398 (0.446)	0.084 (0.247)	0.007 (0.185)	1.349 (0.063)
Shortbelly	*S. jordani*	0.487 (0.146)	1.818 (0.203)	4.387 (0.563)	1.770 (0.242)	0.007 (0.087)	0.443 (0.184)	1.435 (0.056)
Dogfish	*Squalus acanthias*	0.398 (0.134)	0.127 (0.068)	7.650 (0.921)	1.084 (0.122)	0.007 (0.183)	0.075 (0.028)	0.896 (0.012)
Shortspine thornyhead	*Sebastolobus alascanus*	0.899 (0.101)	1.405 (0.083)	5.548 (0.665)	1.126 (0.124)	0.206 (0.059)	0.255 (0.051)	1.056 (0.020)
Widow	*S. entomelas*	0.450 (0.234)	1.294 (0.136)	2.245 (0.381)	0.008 (0.356)	0.007 (0.138)	0.007 (0.123)	0.858 (0.071)
Yelloweye	*S. ruberrimus*	29.173 (7.910)	0.628 (0.377)	6.898 (3.124)	0.609 (0.348)	1.618 (0.487)	0.309 (0.255)	0.943 (0.087)
Yellowtail	*S. flavidus*	0.533 (0.177)	1.360 (0.163)	3.342 (0.503)	1.373 (0.220)	0.007 (0.223)	0.257 (0.204)	1.158 (0.057)

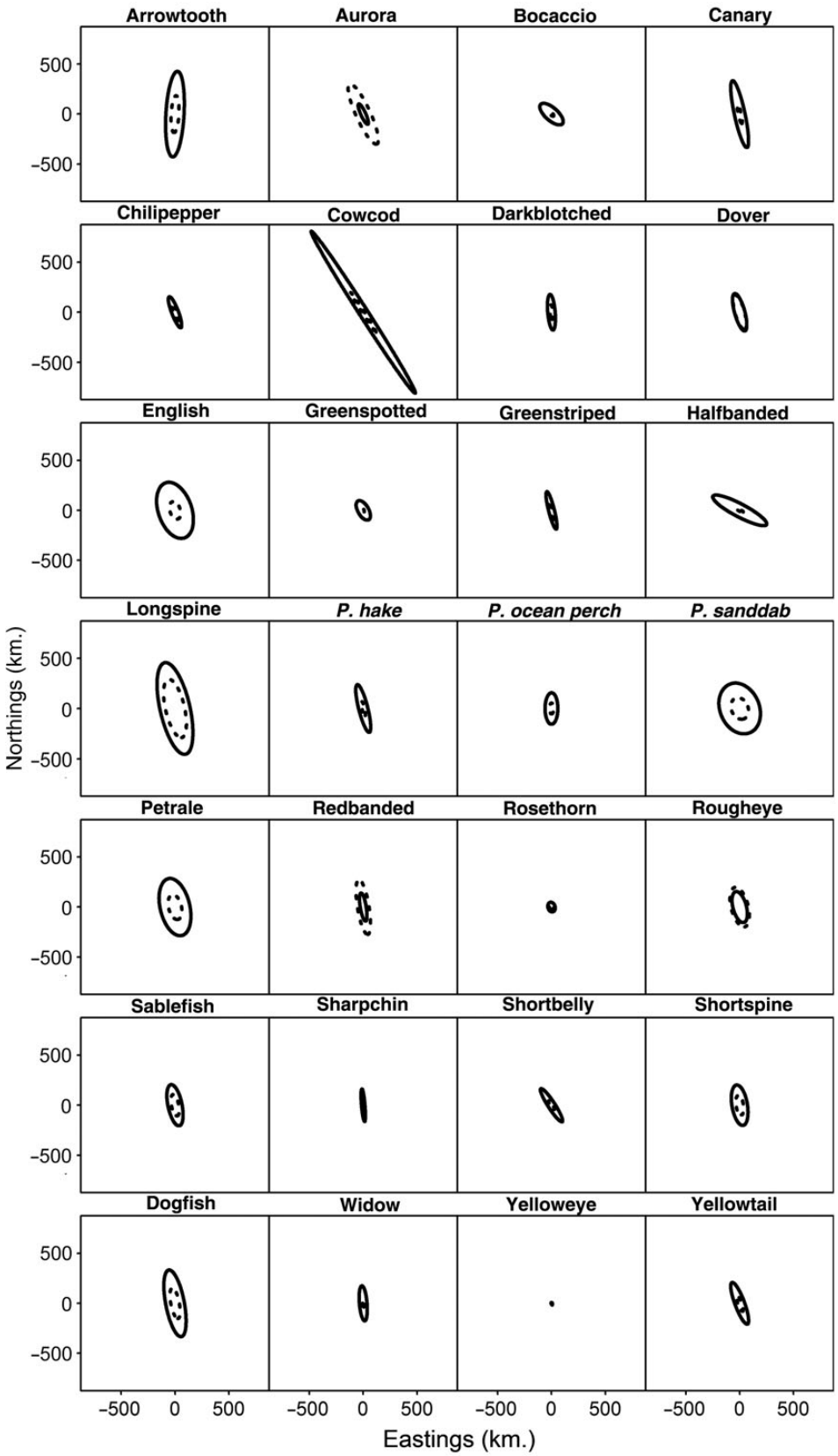

Figure 2. Ellipses representing estimates of geometric anisotropy for each species (dashed line: spatial variation in encounter probability; solid line: spatial variation in positive catch rate), where the line signifies the distance (from a point located at ⟨0,0⟩) where the correlation will have dropped to 10%, e.g. an ellipse that is stretched North – South signifies that densities are correlated over a longer distance moving North – South than East – West.

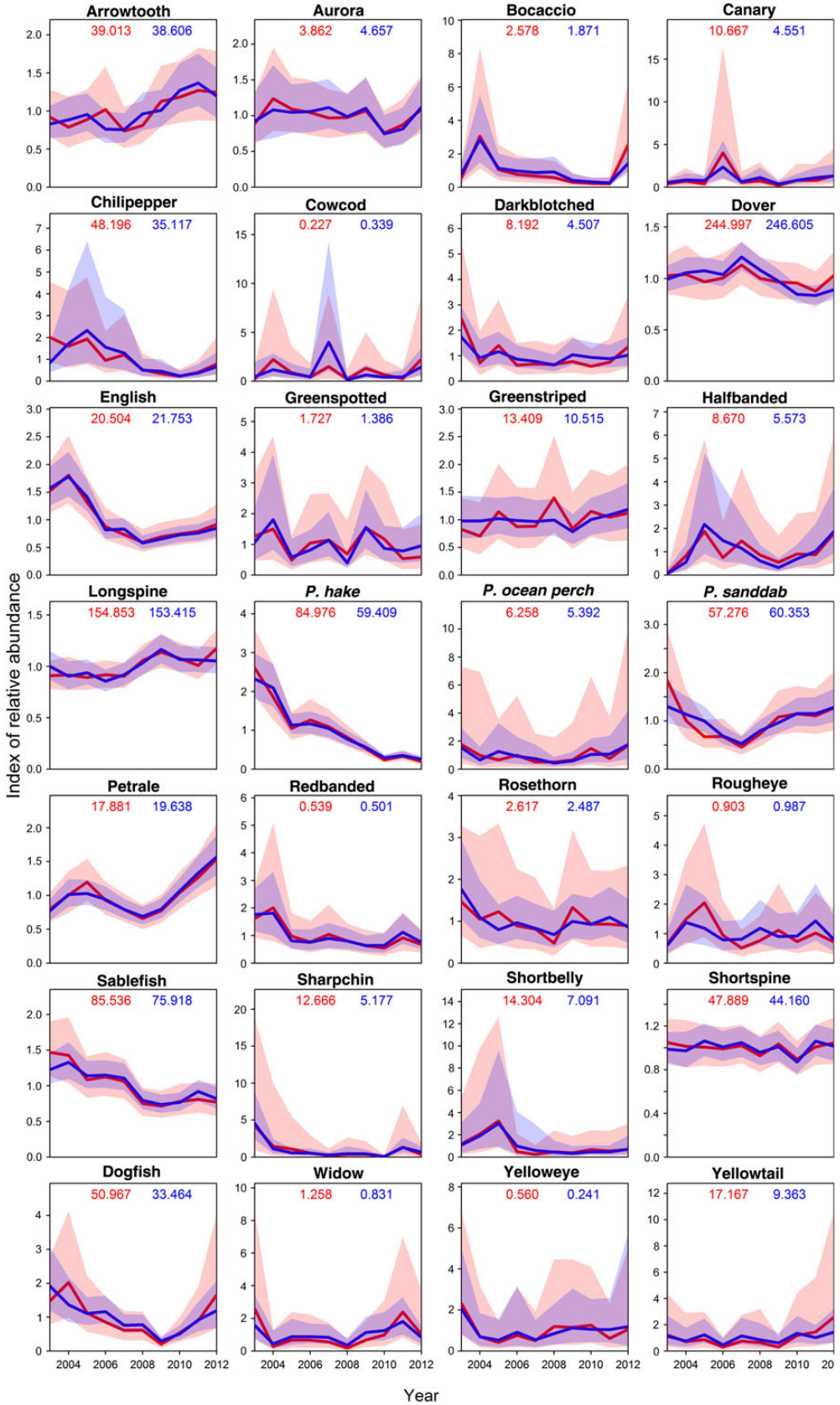

Figure 3. Estimated indices of relative abundance $\hat{l}_t = \hat{b}_t/((1/n_t)\sum_{t=i}^{n_t}\hat{b}_t)$, where $\hat{b}_t$ is estimated abundance (lines) with 95% intervals for the geostatistical delta-GLMM (blue) and the Bayesian stratified delta-GLMM (red; where purple is their overlap), and where the average abundance $(1/n_t)\sum_{t=i}^{n_t}\hat{b}_t$ across the time series is shown at the top of each panel for each model.

The difference between the 20% increase in the simulation study and the 60% increase in the 28 West Coast groundfish could be caused by many factors, including that most groundfish had a lower encounter probability (and hence less information for estimating abundance indices) than the data that we simulated. Regardless of the exact magnitude, we therefore conclude that the geostatistical model uses limited available information more efficiently than the stratified model. This interpretation is consistent with previous spatial analyses (Shelton *et al.*, 2014; Thorson *et al.*, 2015). In general, these analyses conclude that, by explaining spatial variation in densities, geostatistical models can minimize residual (unexplained) variability, and therefore capture a more precise snapshot of biological processes than a spatially stratified model.

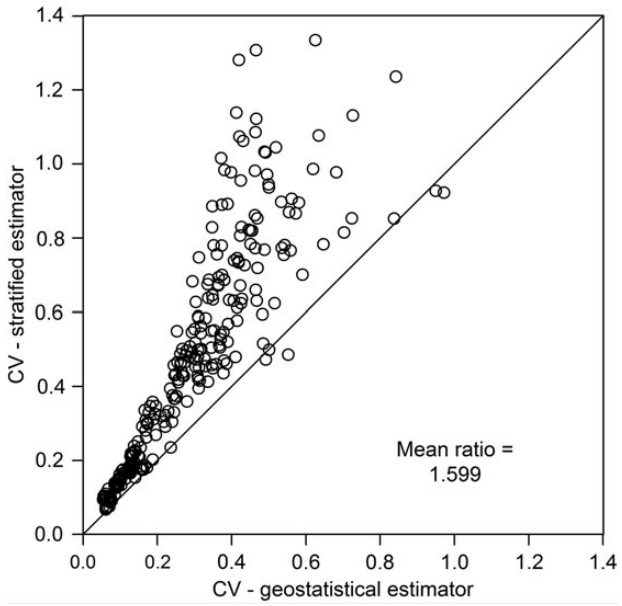

Figure 4. Ratio of coefficient of variation from the geostatistical delta-GLMM (*x*-axis) and Bayesian stratified delta-GLMM (*y*-axis) for each of 28 species and 10 years. The average ratio of CVs is also shown in the bottom-right of the panel (see Eq. (13) and main text for details).

However, we cannot eliminate the possibility that this improved precision occurs at the cost of increased bias relative to a design-based estimator, given that violation of model assumptions can result in bias for any model-based estimator (Smith, 1990). In particular, the Gaussian random fields used to approximate spatial variation imply a particular "prior" on the distribution of spatial variation, and may be biased (in a design-based sense; Smith, 1990) or less statistically efficient for types of spatial variation (e.g. oscillatory behavior) that differ from this prior (see Thorson *et al.*, 2014 for a discussion of priors on functions). The statistical performance and in particular, evidence for estimation bias must therefore be explored via simulation for other fishery contexts. We therefore envision an iterative process, where improved spatial models (e.g. Kristensen *et al.*, 2014) can be used to inform the "states of nature" that are then simulated when evaluating model-based estimators such as the geostatistical model (Punt, 2008).

We also note the importance of ongoing research for generic bias-correction in maximum likelihood mixed-effects models. This issue arises whenever an analyst seeks an unbiased estimator for a derived quantity (e.g. abundance $\hat{b}_t$) that arises as a non-linear function of estimated random effects, in which case a plug-in estimator (as commonly reported by AD Model Builder, Fournier *et al.*, 2012) does not account for uncertainty regarding random effects when reporting the derived quantity. Model-specific bias-correction estimators have been proposed for other contexts (e.g. recruitment deviations in population dynamics models, Methot and Taylor, 2011), although it is not clear how to generalize

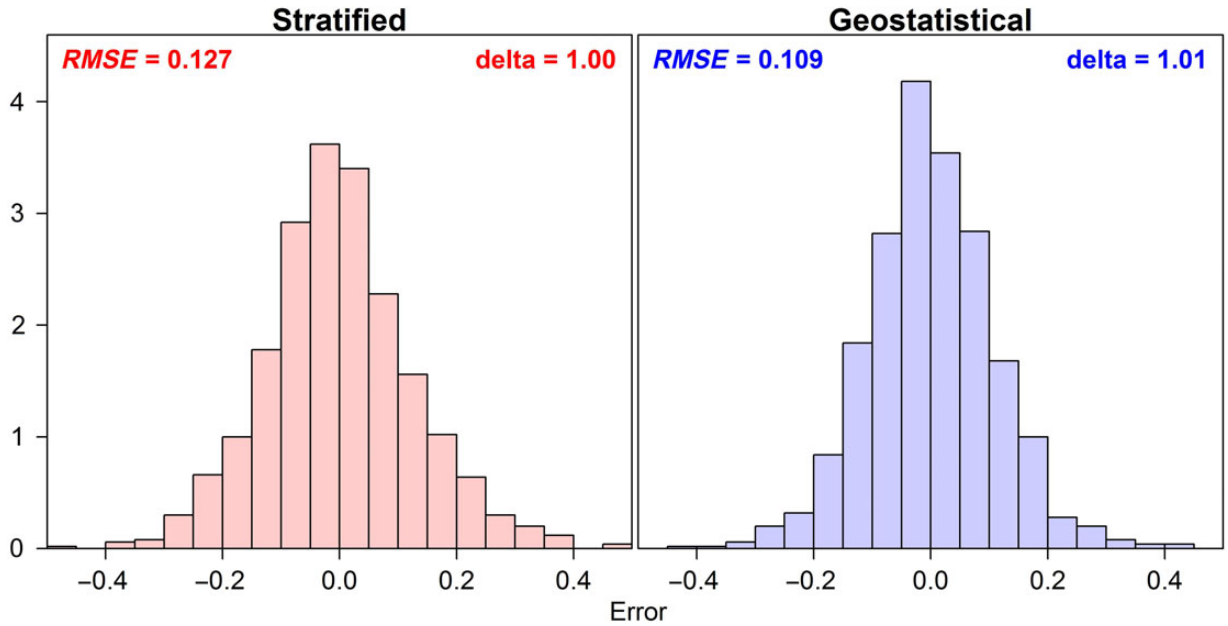

Figure 5. Histogram showing the distribution of error when estimating relative abundance, $\log(\hat{I}_t) - \log(I_t)$ (see main text for details), for 100 replicates of the numerical experiment, where data are generated from a spatially explicit model and are fitted by the stratified delta-GLMM (left panel) or the geostatistical delta-GLMM (right panel).

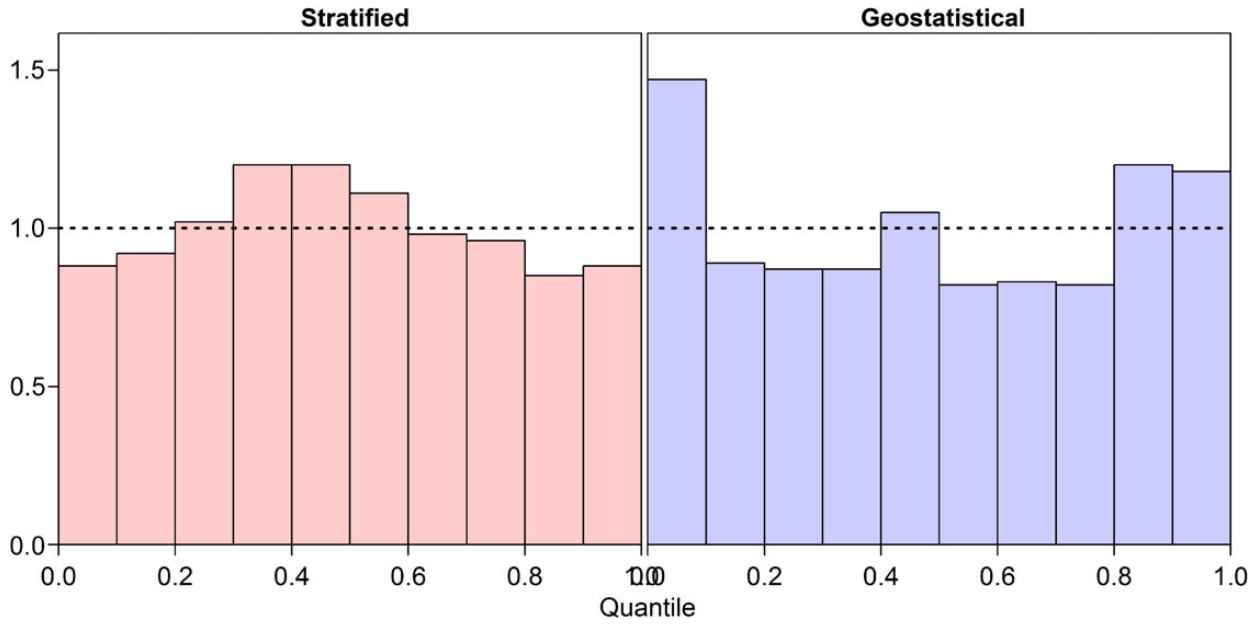

Figure 6. Histogram showing the distribution of quantiles for 100 replicates of the numerical experiment, where data are generated from a spatially explicit model and are fitted by the stratified delta-GLMM (left panel) or the geostatistical delta-GLMM (right panel), where a well-calibrated estimate of intervals will have a uniform distribution (shown as dotted line: see main text for details).

these methods to other model types. K. Kristensen, H. Skaug, and J. Thorson are conducting ongoing research comparing Taylor series expansion and Monte-Carlo simulation with the bias-correction method proposed by Tierney *et al.* (1989), which has been implemented in a development version of template model builder. These comparisons will require further testing, and their implications extend well beyond the geostatistical model developed here. However, the simulation results reported here provide little support for thinking that the plug-in estimator results in biased estimates of abundance trends.

The present study contributes to a growing literature suggesting that geostatistical models are more statistically efficient (i.e. have lower estimation imprecision for a given amount of data) than analysing data using spatially stratified models. This implies that geostatistical estimators may also be appropriate when summarizing other types of data for use in stock assessment. For example, age- and length-composition sampling is usually conducted simultaneously with catch-rate sampling, and compositional data can be "standardized" to estimate the effective sample size for a given data set (Thorson, 2014). Additionally, catches for fish complexes are usually subsampled when estimating species-specific catches (Shelton *et al.*, 2012), and must then be modelled to estimate species composition for unsampled seasons or areas. We hypothesize that the statistical efficiency of these models may be improved by using a geostatistical estimator, i.e. pooling information for nearby port- or creel-samples.

We also note that the current geostatistical estimator calculates total abundance by integrating densities across the spatial domain of the population or survey. It therefore automatically estimates densities in un-sampled areas (using habitat information and nearby samples), and represents a new tool for the "imputation" approach to index standardization (Carruthers *et al.*, 2011; Walters, 2003). Imputation is more commonly discussed for fishery-dependent catch-rate data, where fisher targeting may cause available data to be un-representative of overall population abundance. In the geostatistical literature, this is referred to as "preferential sampling", and it will cause a geostatistical estimator to be biased whenever the sampling intensity is correlated with underlying population densities (Diggle and Ribeiro, 2007). However, future research could explore models that jointly approximate the sampling intensity (i.e. the location of available data) and sampling response (i.e. catch rates at sampled locations), thus presenting a path for modifying the geostatistical approach for fishery-dependent data (see discussion in Diggle and Ribeiro, 2007).

Finally, the geostatistical model developed here can also be used when identifying which habitat variables contribute to observed variation in densities, as well as the cumulative impact of measured (e.g. depth) and unmeasured (e.g. biogenic) variables on realized encounter probabilities and positive catch rates (National Marine Fisheries Service (NMFS), 2013; Shelton *et al.*, 2014). Understanding habitat impacts is important when designating areas that warrant spatial management efforts, and maps of species densities is important when interpreting potential impacts of climate shifts on marine species (Pinsky *et al.*, 2013). We therefore believe that geostatistical methods such as this will grow in importance during the coming decades, both for basic research and tactical fisheries management.

Acknowledgements

We thank K. Ono and B. Feist, who have both contributed to discussions regarding spatial modelling, J. Hastie and A. Berger for comments on an earlier draft, and K. Kristensen for ongoing help and maintenance for template model builder. We also thank the many NOAA scientists who have contributed to the NWFSC Shelf-Slope survey since 2003, and B. Horness for providing access to these data, and two anonymous reviewers for comments on an earlier draft.

References

Bigelow, K. A., and 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., Gardner, B., Maunder, M., Berg, C. W., Brooks, M., Comita, L., Crone, E., *et al.* 2013. Strategies for fitting nonlinear ecological models in R, AD Model Builder, and BUGS. Methods in Ecology and Evolution, 4: 501–512.

Bradburn, M. J., Keller, A. A., and Horness, B. H. 2011. The 2003 to 2008 US West Coast bottom trawl surveys of groundfish resources off Washington, Oregon, and California: estimates of distribution, abundance, length, and age composition. NOAA Technical Memorandum, NMFS-NWFSC-114. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northwest Fisheries Science Center, Seattle, WA.

Burnham, K. P., and Anderson, D. 2002. Model Selection and Multi-Model Inference. Springer, New York.

Carruthers, T. R., Ahrens, R. N., McAllister, M. K., and Walters, C. J. 2011. Integrating imputation and standardization of catch rate data in the calculation of relative abundance indices. Fisheries Research, 109: 157–167.

Cochran, W. G. 1977. Sampling Techniques, 3rd edn. John Wiley & Sons. 428 pp.

Cressie, N., Calder, C. A., Clark, J. S., Hoef, J. M. V., and Wikle, C. K. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecological Applications, 19: 553–570.

Cressie, N., and Wikle, C. K. 2011. Statistics for Spatio-Temporal Data. John Wiley & Sons, Hoboken, NJ.

Diggle, P., and Ribeiro, P. J. 2007. Model-Based Geostatistics. Springer, New York. 242 pp.

Diggle, P. J., Menezes, R., and Su, T. 2010. Geostatistical inference under preferential sampling. Journal of the Royal Statistical Society: Series C (Applied Statistics), 59: 191–232.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A., *et al.* 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software, 27: 1–17.

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

Gertseva, V., and Thorson, J. T. 2013. Status of the darkblotched rockfish resource off the continental U.S. Pacific Coast in 2013. National Marine Fisheries Service, Northwest Fisheries Science Center, Fisheries Resource and Monitoring Division, Seattle, WA.

Helser, T. E., Punt, A. E., and Methot, R. D. 2004. A generalized linear mixed model analysis of a multi-vessel fishery resource survey. Fisheries Research, 70: 251–264.

Kristensen, K., Thygesen, U. H., Andersen, K. H., and Beyer, J. E. 2014. Estimating spatio-temporal dynamics of size-structured populations. Canadian Journal of Fisheries and Aquatic Sciences, 71: 326–336.

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. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73: 423–498.

Lo, N. C., Jacobson, L. D., and Squire, J. L. 1992. Indices of Relative Abundance from Fish Spotter Data based on Delta-Lognormal Models. Canadian Journal of Fisheries and Aquatic Sciences, 49: 2515–2526.

MacCall, A. D. 1990. Dynamic Geography of Marine Fish Populations. University of Washington Press, Seattle, WA.

Magnusson, A., Punt, A. E., and Hilborn, R. 2013. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC. Fish and Fisheries, 14: 325–342.

Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A., Low-Choy, S. J., Tyre, A. J., *et al.* 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecology Letters, 8: 1235–1246.

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

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

Methot, R. D., Tromble, G. R., Lambert, D. M., and Greene, K. E. 2014. Implementing a science-based system for preventing overfishing and guiding sustainable fisheries in the United States. ICES Journal of Marine Science, 71: 183–194.

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

National Marine Fisheries Service (NMFS). 2013. Groundfish essential fish habitat synthesis: a report to the Pacific Fisheries Management Council. NOAA NMFS Northwest Fisheries Science Center, Seattle, WA.

Petitgas, P. 2001. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. Fish and Fisheries, 2: 231–249.

Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., and Levin, S. A. 2013. Marine taxa track local climate velocities. Science, 341: 1239–1242.

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). Vienna, Austria.

Punt, A. E. 2008. Refocusing stock assessment in support of policy evaluation. *In* Fisheries for Global Welfare and Environment, pp. 139–152. Ed. by K. Tsukamoto, T. Kawamura, T. Takeuchi, T. D. Beard, and M. J. Kaiser. TerraPub, Tokyo.

Rasmussen, C. E., and Williams, C. K. I. 2006. Gaussian Processes for Machine Learning. MIT press, Cambridge, MA.

R Core Development Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

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

Schnute, J. T., Boers, N., Haigh, R., Grandin, C., Johnson, A., Wessel, P., and Antonio, F. 2013. PBSmapping: Mapping Fisheries Data and Spatial Analysis Tools. http://CRAN.R-project.org/package=PBSmapping.

Searle, S. R., Casella, G., and McCulloch, C. E. 1992. Variance Components. John Wiley & Sons, Hoboken, New Jersey. 536 pp.

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. Canadian Journal of Fisheries and Aquatic Sciences, 69: 231–246.

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

Skaug, H., and Fournier, D. 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. Computational Statistics & Data Analysis, 51: 699–709.

Smith, S. J. 1990. Use of statistical models for the estimation of abundance from groundfish trawl survey data. Canadian Journal of Fisheries and Aquatic Sciences, 47: 894–903.

Stefansson, G. 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES Journal of Marine Science, 53: 577–588.

Stephens, A., and MacCall, A. 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fisheries Research, 70: 299–310.

Thorson, J. T. 2014. Standardizing compositional data for stock assessment. ICES Journal of Marine Science: Journal du Conseil, 71: 1117–1128.

Thorson, J. T., Elizabeth Clarke, M., Stewart, I. J., and Punt, A. E. 2013. The implications of spatially varying catchability on bottom trawl surveys of fish abundance: a proposed solution involving underwater vehicles. Canadian Journal of Fisheries and Aquatic Sciences, 70: 294–306.

Thorson, J. T., Hicks, A. C., and Methot, R. 2015. Random effect estimation of time-varying factors in stock synthesis. ICES Journal of Marine Science, 72: 178–185.

Thorson, J. T., and Minto, C. 2014. Mixed effects: a unifying framework for modelling in aquatic ecology. ICES Journal of Marine Science, 71: 296–298.

Thorson, J. T., Ono, K., and Munch, S. B. 2014. A Bayesian approach to identifying and compensating for model misspecification in population models. Ecology, 95: 329–341.

Thorson, J. T., Skaug, H., Kristensen, K., Shelton, A. O., Ward, E. J., Harms, J., and Benante, J. in press-b. The importance of spatial models for estimating the strength of density dependence. Ecology. doi: http://dx.doi.org/10.1890/14-0739.1.

Thorson, J. T., and Ward, E. 2013. Accounting for space-time interactions in index standardization models. Fisheries Research, 147: 426–433.

Thorson, J. T., and Ward, E. J. 2014. Accounting for vessel effects when standardizing catch rates from cooperative surveys. Fisheries Research, 155: 168–176.

Tierney, L., Kass, R. E., and Kadane, J. B. 1989. Fully exponential Laplace approximations to expectations and variances of nonpositive functions. Journal of the American Statistical Association, 84: 710–716.

Walters, C. 2003. Folly and fantasy in the analysis of spatial catch rate data. Canadian Journal of Fisheries and Aquatic Sciences, 60: 1433–1436.

Wilberg, M. J., Thorson, J. T., Linton, B. C., and Berkson, J. 2010. Incorporating time-varying catchability into population dynamic stock assessment models. Reviews in Fisheries Science, 18: 7–24.

Yu, H., Jiao, Y., and Carstensen, L. W. 2013. Performance comparison between spatial interpolation and GLM/GAM in estimating relative abundance indices through a simulation study. Fisheries Research, 147: 186–195.

Handling editor: Mark Maunder