LIMNOLOGY
and
OCEANOGRAPHY: METHODS

ASLO

Limnol. Oceanogr.: Methods 15, 2017, 480–494

doi: 10.1002/lom3.10174

A comparison of methods to estimate abundance and biomass from belt transect surveys

Jui-Han Chang, Burton V. Shank, Deborah R. Hart*
Northeast Fisheries Science Center, Woods Hole, Massachusetts

Abstract

It is becoming increasingly popular to use continuously collected acoustic or optical data to estimate abundance or biomass of fish and invertebrates. However, data from such systems are typically highly spatially autocorrelated and zero-inflated, and thus simple design-based estimation techniques are not applicable. Model-based estimation methods can be used to extrapolate observations along the observed track to larger areas. We tested the precision and accuracy of three model-based methods using both simulations and field data: Ordinary kriging (OK), Generalized Additive Models with kriged model residuals (GAM + OK), and Generalized Additive Mixed Models with kriged model residuals (GAMM + OK), along with a design-based method (stratified mean, SM). The GAMM + OK method treats small-scale variations as random effects, whereas the other approaches aggregate nearby data to reduce autocorrelation and random errors. We found that the GAM + OK method with relatively small aggregation lengths generally gave the best performance of the model-based methods in terms of both accuracy and precision, followed by GAMM + OK. SM estimates were more accurate and precise than the model-based estimates in the simulations, but only when the study region was stratified accurately. Based on the simulation and field data analysis results, we selected the GAM + OK method to estimate scallop abundance and biomass for the Georges Bank and the Mid-Atlantic Bight regions for the years 2011–2015. We also provided SM estimates based on careful stratifications to validate the model-based estimates.

Many types of survey data are collected continuously, such as acoustic data and photographs from towed or autonomous underwater vehicles. Because the samples produced from these "belt transect" surveys are not random, simple design-based estimators of the population mean and variance are not directly applicable. For this reason, model-based estimation is often used to estimate population abundances (Petitgas 1993; Simard et al. 1993; Maravelias et al. 1996; Simard and Lavoie 1999; Páramo and Roa 2003; Hedley and Buckland 2004; Mello and Rose 2005; Georgakarakos and Kitsiou 2008; Rivoirard et al. 2008; Williams et al. 2011) although design-based methods have also been used (Jolly and Hampton 1990; Brandt et al. 1991; Singh et al. 2013).

The fundamental difference between design-based and model-based approaches is that for the design-based methods, the population is regarded as fixed and the survey data are the measured characteristics of this population, whereas for the model-based methods, the observed population is only one realization of a stochastic process (Särndal et al. 1978; Smith 1990). Design-based methods based on stratifications require no assumptions regarding the underlying population, but the samples drawn from each stratum must be randomized so that they are independent and identically distributed. Failure to meet the randomness requirement may cause bias in variance estimates (Cochran 1977). On the other hand, although model-based estimation methods do not require random sampling, they are typically based on strong assumptions regarding the nature of the underlying population. For example, one popular model-based method, ordinary kriging (OK), is based on the assumptions that the population means and covariances are spatially stationary. Real populations may not satisfy these assumptions, which can cause the model-based estimators to be biased.

The purpose of this paper is to evaluate a number of potential methods to estimate the population abundance or biomass of non-stationary populations, based on surveys that collect abundance and biomass data continuously, using both simulations and real data. These include OK, and two variations of regression kriging (RK) that can take into account large-scale trends and covariates in the data: Generalized Additive Models on spatially aggregated data with kriged model residuals (GAM + OK), and Generalized Additive Mixed Models where small-scale variations are treated as random effects, combined with kriged model residuals

*Correspondence: deborah.hart@noaa.gov

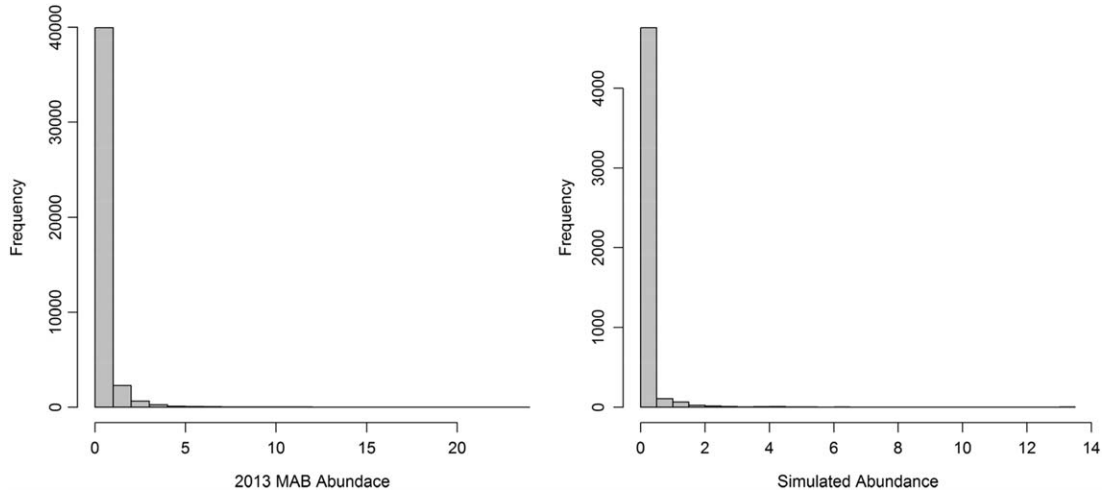

Fig. 1. Histogram of HabCam abundance data (numbers/m^2) for MAB in 2013 (left) and from one of the simulated populations (right).

(GAMM + OK). In addition, a design-based method based on stratified means (SM) is also evaluated.

Although the methods described here should be applicable to a wide variety of continuously collected optical and acoustic data, we focus here on observations of sea scallops (*Placopecten magellanicus*) collected using the vessel-towed underwater digital habitat mapping camera system (HabCam) as an example case. HabCam was developed through collaborations between scientists at the Woods Hole Oceanographic Institution, the Northeast Fisheries Science Center (NEFSC), and with commercial fishermen to survey benthic communities, and to map sea floor habitats (Howland et al. 2006; Taylor et al. 2008; NEFSC 2010). The cameras on HabCam take rapid-fire still photos of the sea floor (typically 6/s) as it is towed at speeds between 5 and 7 knots at about 2 m above the bottom. Region-scale HabCam surveys for sea scallops were conducted on Georges Bank (GB) in 2011, and on both GB and the Mid-Atlantic Bight (MAB) in 2012–2015. Scallop data from HabCam are highly spatially autocorrelated and zero-inflated (i.e., a high percentage of the data are zeros; Table 1; Fig. 1), reflecting the patchiness of scallop distributions and the continuous nature of the observations.

Materials and procedures

Simulation design

The simulated area was 50 km longitude and 100 km latitude (the shape and size are similar to Hudson Canyon South rotational management area in the MAB; NEFSC 2014) with a 100 m grid size. Scallop spatial distributions are nonstationary due to the influences of the physical and biological environment such as substrate, depth, temperature, and predator distributions (Brand 2006; Hart 2006). The simulated scallop populations are therefore assumed to vary nonrandomly according to large-scale trends, termed the "first-order effect." For simplicity in the simulations, these trends are assumed to be (non-linear) functions of longitude only. In reality, depth and other environmental factors may be important predictors of the trend; longitude is treated as a surrogate for depth and other environmental factors in the simulations. Stationary "second-order effects," representing small-scale spatially autocorrelated variability, were added to the first-order trends. Various first-order and second-order effects were simulated to test whether the abundance and biomass estimation methods are robust to type of spatial distributions of the underlying population.

We simulated the first-order trend using a double logistic function

$$p_{i,j}=\begin{cases}\dfrac{1}{1+\exp\left(-a\ (i-b)\right)} & i \le \dfrac{i_{\max}}{2} \\ \dfrac{1}{1+\exp\left(a\ \left(i-b-\dfrac{i_{\max}}{2}\right)\right)} & i > \dfrac{i_{\max}}{2}\end{cases}, \quad (1)$$

where a and b parameters determine the shape of the logistic curve, i and j are the longitude and latitude, respectively, and $i_{\max}$ is the east boundary of the longitudes. The simulated first-order effects are greatest in the middle and decrease logistically toward the left and right edge of the simulation domain (Fig. 2); this pattern mimics the observed distribution of sea scallops in the MAB, where scallop densities are the greatest at intermediate depths (Hart 2006). We simulated two types of first-order effects: one where the population is more concentrated in the middle area, whereas they are more spread out in the second (Fig. 2).

We simulated the second-order effects using stationary Gaussian random fields with spherical isotropic covariance structures (Cressie 1993):

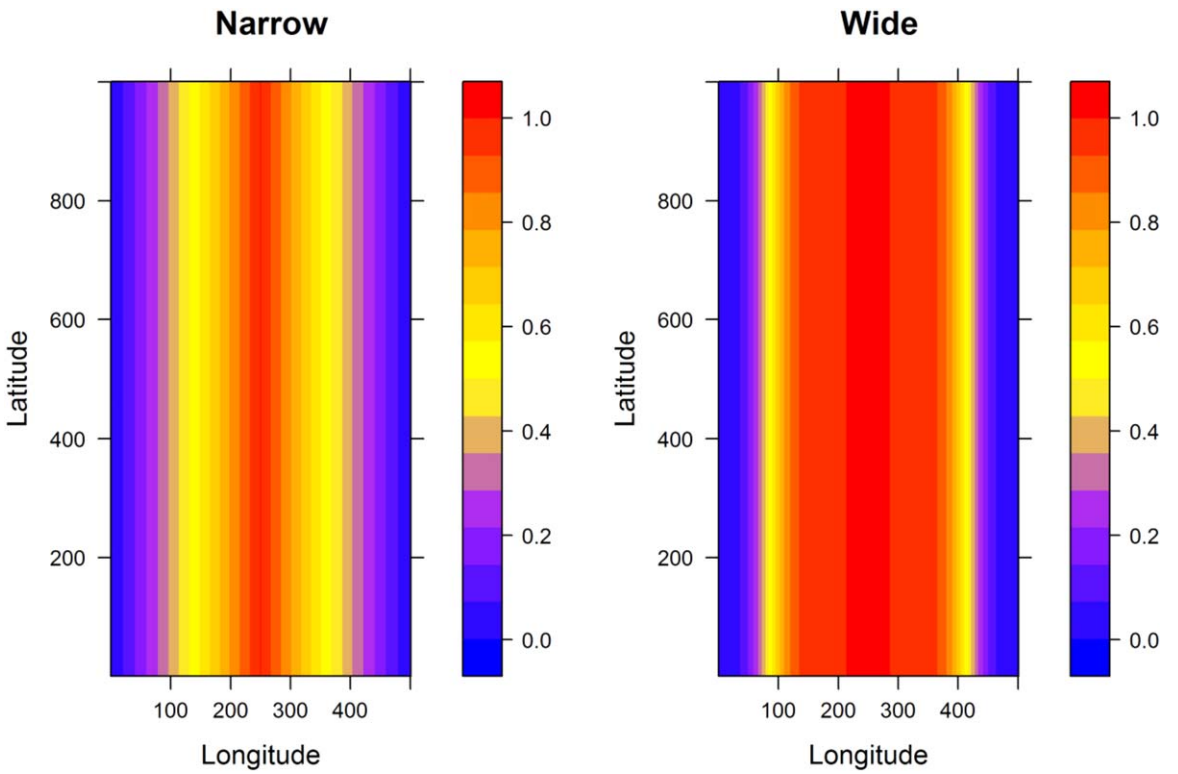

Fig. 2. The two types of first-order effects used to simulate scallop populations: one where the high density central region is narrow (left) and one where it is wide (right). The colors represent simulated first-order effects to scale scallop densities.

$$\gamma(h)=\begin{cases} 0 & h=0 \\ c_0+c_1\left\{\frac{3}{2}\frac{h}{r}-\frac{1}{2}\left(\frac{h}{r}\right)^3\right\} & 0<h\leqq r \\ c_0+c_1 & h\geqq r, \end{cases} \quad (2)$$

where c_0, c_1, and r are the nugget, partial sill, and range parameter, respectively, and h is the distance between two points. The nugget/sill ratio ($\frac{c_0}{c_0+c_1}$) determines the extent of the random variability and r determines the aggregation size of the second-order effects. We simulated combinations of two levels of the nugget/sill ratio (0.4 and 0.6) and two levels of the range parameter (100 and 300) resulting in four types of second-order effects: small or large aggregation sizes, with low or high random noise (Fig. 3). Parameter values were based on estimates from the actual HabCam data.

Scallop distributions are patchy, resulting in data that are highly zero-inflated (Table 1; Fig. 1). To reflect the extent of the zero inflation observed in the actual data, only 10% of the locations with the highest sums of first- and second-order effects in the simulations were taken to have non-zero scallop densities; the other 90% of the sites were set to zero density (Fig. 1).

Eight types of population distributions, from two types of first-order and four types of second-order effects were simulated (Fig. 4). We generated 30 realizations for each population type, and then scaled the total abundance and biomass of each realization so that total biomass and abundance was the same. Each simulated population was surveyed using 30 different tracks (where the starting point and first turn of the track were varied). Shape and direction of the simulated tracks was designed to mimic the actual HabCam survey design, in which the long transects are approximately in the direction of the gradient of density. The length of these long transects are alternated: one long transect extends to the boundary of the survey area, followed by a short transect extending to the edge of the middle high density area (NEFSC 2014). This design covers the middle higher density areas more intensely than the more marginal areas toward the edges of the domain in order to improve survey efficiency. Additionally, it has cross-transects near the high density middle portion of the domain that facilitate estimation of anisotropy. By contrast, a simpler design where each main transect was the same length would have all its cross transects at the edges of the domain, where densities are close to zero, which would give less information on the directional structure of the population.

We used model-based and designed-based methods to estimate total biomass and abundance for the simulated populations. These methods were evaluated using relative bias (RBias) and relative root mean square error (RRMSE)

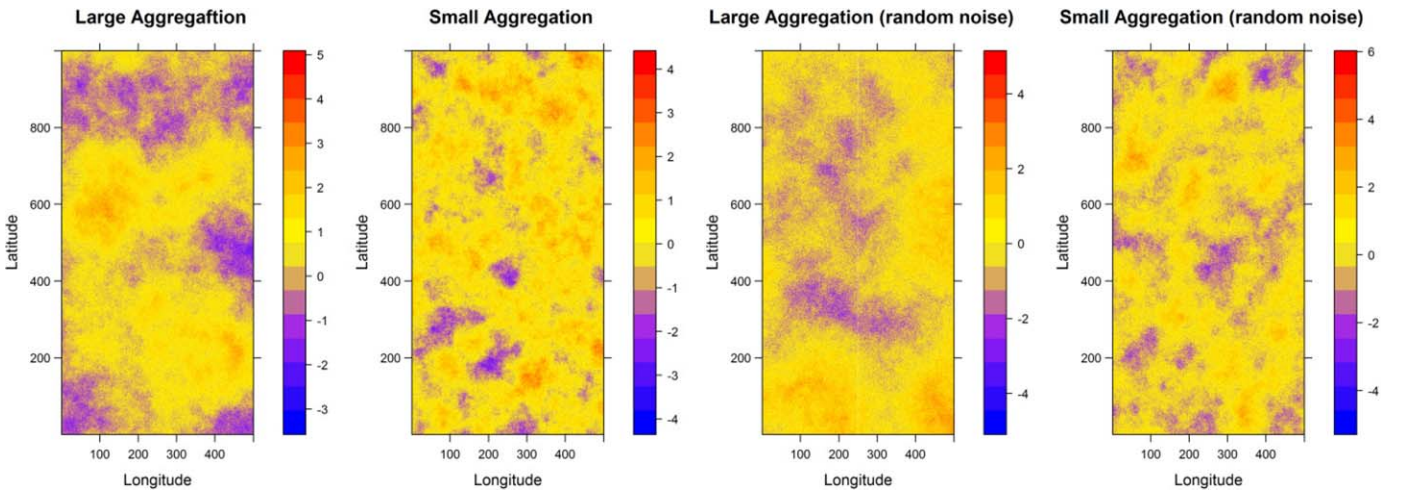

Fig. 3. The four types of second-order effects used to simulate scallop populations: large and small aggregations with low random noise and large and small aggregations with high random noise (from left to right). The colors represent simulated scallop densities based on second-order effects only.

$$\text{RBias} = \frac{\frac{\sum_{i=1}^{n}(\hat{T}_i - \mu)}{n}}{\mu} \tag{3}$$

$$\text{RRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(\hat{T}_i - \mu)^2}{n}}}{\mu}, \tag{4}$$

where $\hat{T}_i$ is the estimated total biomass or abundance for sample sets i, μ is the true population abundance or biomass, and n is the total number of sample sets analyzed. We also evaluated the RBias of the estimated coefficient of variation (CV) of the abundance or biomass estimates.

Model-based estimation

Kriging is one of the most widely used geostatistical method for spatial interpolation (Webster and Oliver 2001). We tested performance of the three different kriging methods: OK, GAM + OK, and GAMM + OK on the simulated scallop populations. OK is a standard kriging method based on the assumptions of stationary means and covariances (Webster and Oliver 2001; Hengl 2009). In some cases, the population may be anisotropic, that is, its variability may be directionally dependent. Based on the assigned first-order effects, the simulated populations should have the largest variations along the horizontal axis, due to the strong longitudinal (depth) effects. Therefore, we built both the isotropic and anisotropic models (0° and 90°) and tested four types of commonly used variogram models including spherical, exponential, Gaussian, and Matérn models (Cressie 1993). Of these models, the one that minimized the root mean square error (RMSE, square root of sum of squared deviations of the model predictions from the observed values) was selected. Total abundance or biomass (T) and its variance from this model were calculated as:

Table 1. Summary of HabCam data for sea scallops. "Annotated images" is the total number of images where scallops were noted and measured. "Images w/scallops" is the number of annotated images in which scallops were observed.

Stock	Year	Annotated images	Images w/scallops
GB	2011	202,257	21,428
GB	2012	36,304	7189
GB	2013	33,864	4671
GB	2014	47,452	7107
GB	2015	50,558	6025
MAB	2012	20,969	2095
MAB	2013	42,213	3627
MAB	2014	45,393	5997
MAB	2015	61,771	10,000

$$\hat{T} = A\sum_{i=1}^{n}\hat{z}_i \tag{5}$$

$$\text{Var}\left(\hat{T}\right) = A^2\sum_{i=1}^{n}\sum_{j=1}^{n}\text{Cov}\left(\hat{z}_i, \hat{z}_j\right), \tag{6}$$

where $\hat{z}_i$ is the kriging estimate at location i and A is the grid size.

RK extends OK to account for a potentially non-linear global trend. This trend can be estimated using a generalized regression model (e.g., GLM or GAM), potentially with a series of ancillary predictors, and then OK is performed on the residuals of the regression model to model the second-order effects (Odeh et al. 1995; Hengl 2009). The final RK predictions are obtained by summing the regression predicted values and the kriged residuals.

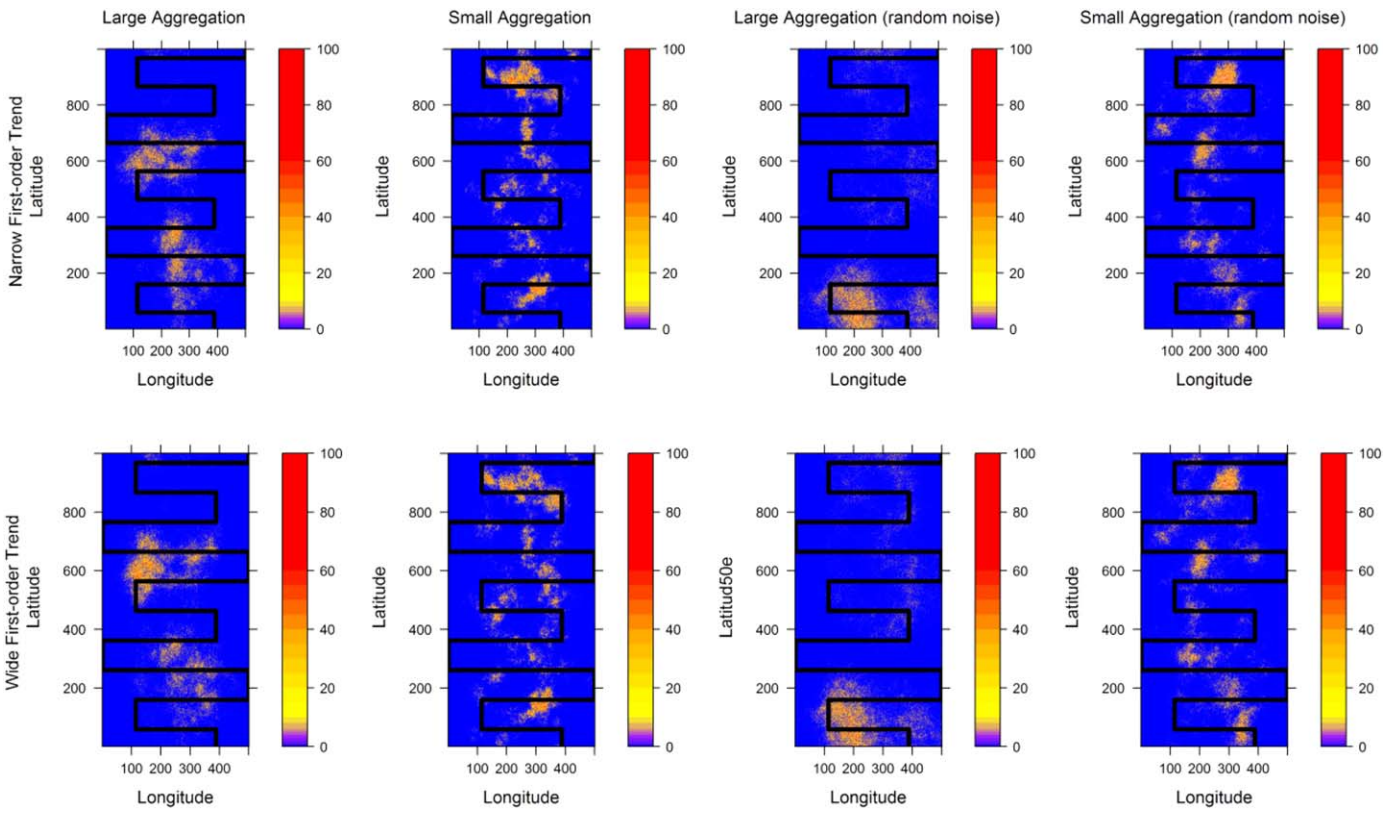

Fig. 4. Example realizations of the eight types of simulated scallop population distribution with an overlaid sampling track (black line). The colors represent simulated scallop biomass densities (g/m^2).

Because HabCam scallop data are zero-inflated (as is common with population counts), we used a two-staged "hurdle" model to estimate the first-order effects, where presence/absence and the level of biomass or abundance at a non-zero site are modeled separately and then combined to derive the final estimates (Barry and Welsh 2002; Smith et al. 2012; Zuur et al. 2012). The hurdle model is given by the distribution:

$$y=\begin{cases}0 & \text{with probability } p\\ \varphi(\theta) & \text{with probability } 1-p,\end{cases}$$

where p is the probability of a zero observation and θ is a vector of parameters for the statistical distribution φ of positives (abundance or biomass; Smith et al. 2012). A quasi-binomial quasi-likelihood was assumed for the probability of presence model, and a quasi-Poisson quasi-likelihood for the positive model that estimates abundance (or biomass), given presence.

A "quasi-likelihood" is an assumption of the relationship between the variance and mean of the observations (Wedderburn 1974); in the quasi-Poisson, the variance is assumed to be proportional to the mean, whereas in the quasi-binomial, the variance is proportional to p(1-p), where p is the mean. Quasi-likelihoods allow fitting of the GAM(M)s without assuming a specific probability distribution. In particular, they can account for overdispersion, where the variance increases faster than the mean, which is commonly observed in aquatic populations in general, and scallops in particular.

We estimated the first-order effects using a two-dimensional spline function of latitude and longitude in both the GAM and GAMM models. The spatial residuals obtained from the large-scale model were used to estimate fine-scale spatial patterns using OK with the same estimation process described above. We estimated the total abundance and biomass of GAM + OK and GAMM + OK model estimates as:

$$\hat{T}=A\sum_{i=1}^{n}\hat{x}_i\hat{y}_i+\hat{z}_i, \tag{7}$$

where $\hat{x}_i$ is the probability of presence estimate, $\hat{y}_i$ is the estimate of abundance or biomass given that scallops are present, and $\hat{z}_i$ is the kriged residual at location i. By assuming that $\hat{x}$ and $\hat{y}$ are independent, the variance of the $\hat{T}$ was calculated as

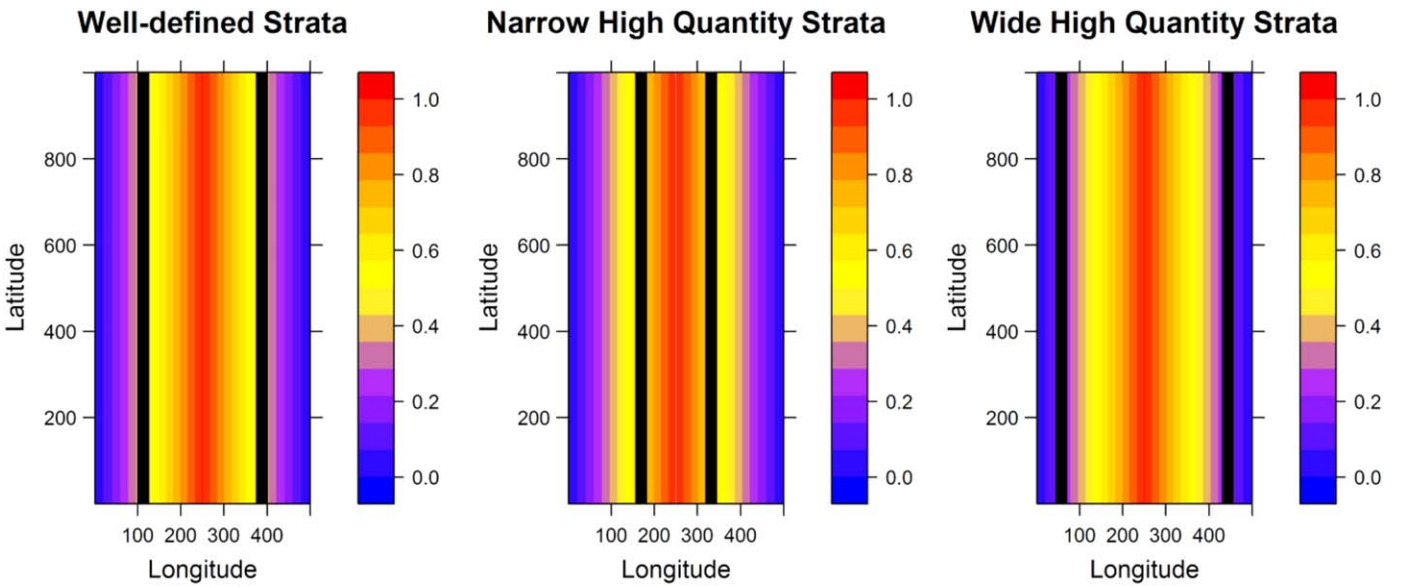

Fig. 5. Alternative types of stratifications (black line) used for SM estimations. The colors represent the simulated scallop densities based on the first-order effects only.

$$\mathrm{Var}\left(\hat{T}\right)=A^2\left[\sum_{i=1}^{n}\left(E^2(\hat{x}_i)\mathrm{Var}(\hat{y}_i)+E^2(\hat{y}_i)\mathrm{Var}(\hat{x}_i)+\mathrm{Var}(\hat{x}_i)\mathrm{Var}(\hat{y}_i)\right)+\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(\hat{z}_i,\hat{z}_j)\right] \tag{8}$$

Data for model-based estimation

In order to reduce the extent of the zero inflation and autocorrelation among nearby data points, the data were blocked into segments of a fixed length along the tracks. For OK and the hurdle GAM, data within each segment were aggregated into a single data point, with its position taken to be the average of the locations of images, weighted by the field of view of the images in that segment. The hurdle GAMM, by contrast, uses each data point individually, but treats within segment variations as random effects. Treating the data within each segment in one of these ways is necessary because nearby data are autocorrelated, which causes the effective sample size to be well less than the total number of data points. In particular, GAMs (without random effects) are based on the assumption that data points are independent, which would be strongly violated if the data were not aggregated. This is a common technique used when analyzing fisheries acoustic data (Mello and Rose 2005). The length of the segment should be sufficient to reduce the degree of random variability and spatial autocorrelation of the data, while at the same time small enough to preserve spatial structures (Mello and Rose 2005). There is no prior knowledge on what the segment length should be used for Atlantic sea scallop and also how sensitive the segment length is for this type of analysis. Therefore, we evaluated the effects of segment length to average the data or determine random effects along the tracks. Scallop aggregations tend to occur at scales of around 1 km (NEFSC 2010), so we tested three segment lengths, 0.75 km, 1.5 km, and 2.25 km. The segment lengths used in the analysis is equivalent to the grid size A, which is the grid size for interpolation.

Design-based estimation

We tested a SM method to estimate the total abundance and biomass from the simulated data. Only the horizontal transects (along lines of latitude in the simulations) were used in the SM estimation. Because some transects do not extend to the low density edges of the domain, while others do, it is necessary to post-stratify the horizontal transects into two strata based on high and low first-order effects (Fig. 5). We calculated the mean and its variance of the simulated scallops (t) by segment (j) and stratum (i) as:

$$\bar{t}_{i,j}=\frac{\sum_{k=1}^{n_{i,j}} t_{i,j,k}}{n_{i,j}} \tag{9}$$

$$\mathrm{Var}\left(\bar{t}_{i,j}\right)=\frac{\mathrm{Var}\left(t_{i,j,k}\right)}{n_{i,j}}, \tag{10}$$

where $n_{i,j}$ is the number of images by segment and stratum. Total abundance (or biomass) and its variance were estimated as (Cochran 1977):

$$\hat{T}=A\sum_{i=1}^{2} S_i\frac{\sum_{j=1}^{n_i}\bar{t}_{i,j}}{n_i} \tag{11}$$

$$\mathrm{Var}\left(\hat{T}\right)=A^2\sum_{i=1}^{2}S_i^2\sum_{j=1}^{n_i}\frac{\mathrm{Var}\left(\bar{t}_{i,j}\right)}{n_i^2}, \tag{12}$$

where n_i is the number of segments by stratum i, and S_i is the area size of stratum i.

The simulation domain was well-stratified based on the first-order trend and the length of the short transects. However, the same precise information may not be available when dealing with the real data. For this reason, we tested the sensitivity of SM estimates to post-stratification error by widening (SMW) and narrowing (SMN) the central high abundance or biomass stratum by 20% (Fig. 5) and then estimated the SMs based on these less perfect stratifications.

HabCam data

The HabCam data were collected during 2011–2015 in GB and 2012–2015 in MAB. We divided the GB and MAB stock regions into 14 subregions based on geographic characteristics and management areas and analyzed them separately because their topology, orientation, and covariance structures differ. Images taken at altitude higher than 4 m were excluded from the analysis because of their poor image qualities. Only scallops with measured shell height larger than 40 mm were used in the analysis because of concerns about full delectability of very small scallops. We converted the shell height (SH) measures to meat weights (g) (MW) using

$$\begin{aligned}\mathrm{MAB}:\ \mathrm{MW}=&-16.88+4.64\log(\mathrm{SH})+1.57\log(D)\\&-0.43\log(\mathrm{SH})\log(D)\end{aligned} \tag{13}$$

$$\mathrm{GB}:\ \mathrm{MW}=\ 14.38+2.826\log(\mathrm{SH})-0.529\log(D)-5.98\log(L), \tag{14}$$

where D is depth (in meters) and L is latitude (Hennen and Hart 2012). Total count and weight in an image were standardized into abundance and biomass per m^2 by dividing by the field of view of the image. A summary of the HabCam data used by year is listed in Table 1.

For model-based estimations on the real data, we enlarged each subregion by 1 km and used the data within this expanded area to build the subregional models. The average of weight or count (t) by image (j) and segment (i) weighted by field of view (f) for every segment along the tracks was calculated as:

$$\bar{t}_i=\sum_{j=1}^{n_i}\frac{f_{i,j}t_{i,j}}{\sum_{j=1}^{n_i}f_{i,j}} \tag{15}$$

The $\bar{t}_i$ was weighted by both variation (s) and number of images (n) in hurdle GAM using

$$w_i=\frac{s_i-s_{(1)}}{2\left(s_{(n_i)}-s_{(1)}\right)}+\frac{n_i-n_{(1)}}{2\left(n_{(n_i)}-n_{(1)}\right)}, \tag{16}$$

where numbers in parentheses represent order statistics (e.g., $s_{(1)}$ represents the 1st order statistic, i.e., the minimum of the s_is). Hurdle GAMs and GAMMs were fitted using quasi-binomial quasi-likelihoods for the presence/absence model and quasi-Poisson quasi-likelihoods for the positive model to estimate the first-order trend with respect to latitude, longitude, and depth. Depth is correlated with latitude and/or longitude. To prevent potential problems cause by this collinearity, latitude and longitude were transformed into composite variables: latitude plus longitude, half of the latitude or longitude plus longitude/latitude. We built models including depth plus one of either latitude, longitude, or a latitude/longitude combination. Depth was included in all of the candidate models because it is one of the most important variables that affect scallop distributions. The maximum amount of knots for each term in the GAM and GAMM was limited to 15 for the interaction terms (reduced to 10 for some of the subregions) and 10 for the single terms to prevent from overfitting. The final first-order model was selected using RMSE from a 10-fold cross validation. We then performed OK on the model residuals, tested isotropic and a series of anisotropic (from 0 to 180 by 20°) residual OK models, and selected the final OK model using the MedSE:

$$\mathrm{MedSE}=\sum_{i=1}^{n}\mathrm{Median}\left(\hat{\bar{t}}_i-\bar{t}_i\right) \tag{17}$$

The 2013 HabCam data was used to evaluate the performance of the three model-based methods on actual data, using a range of segment lengths (0.5–1.75 km by 0.25 km) for estimating the total biomass. The first-order effects were estimated using a smooth function of depth and the selected latitude/longitude combination for both the GAM and GAMM approaches. Model performance was evaluated by comparing model predictions to observations from other surveys that are not used in the estimation model, including dredge surveys from the NEFSC (Hart and Rago 2006) and the Virginia Institute of Marine Sciences, and video drop camera surveys from the School for Marine Science and Technology at the University of Massachusetts, Dartmouth (Stokesbury et al. 2004). Dredge data were expanded using dredge efficiencies of 0.41 on sand substrates and 0.27 on rougher gravel/cobble substrates (NEFSC 2014). Stations from these other surveys were not typically located on the HabCam transects, and thus model interpolations from the HabCam data at the station locations could be compared to the actual survey data. We also used out-of-sample HabCam data, typically center lines that were not part of the basic survey design and were not used to estimate the models, for the same purpose. The MedSE criterion was used to determine the model that best predicts these other survey observations.

For the SM on the real data, each transect was split into segments and the data within each segment were aggregated to help mitigate autocorrelation. We first separated the transects into segments at locations where the direction of

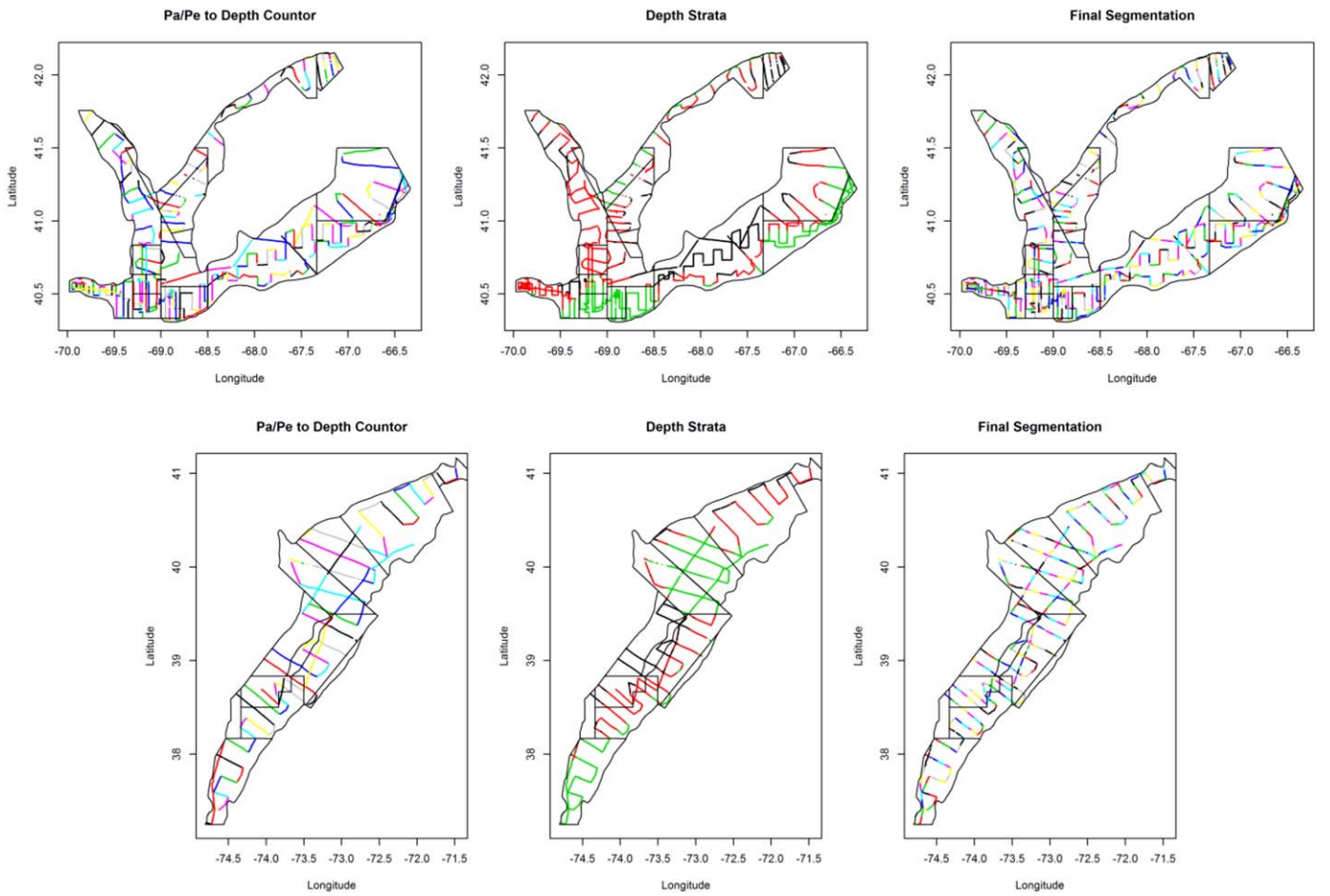

Fig. 6. Transect segmentation for 2015 SM biomass estimation based on orientation to depth contour and distance between points (2 km) (left), depth strata (center), and the final segmentations (right) for GB (upper panel) and MAB (lower panel).

the transects changed between parallel and perpendicular to the depth contour. These were further separated by depth strata or by locations where the distance of any two points in the segment was larger than 2 km. These were broken into even smaller pieces if the length of the segments were longer than 10 km (*see* Fig. 6 for an example).

We estimated thresholds for the depth strata from a maximum likelihood based change-point analysis (Killick et al. 2010), using the depth partial residuals from GAMs of abundance or biomass with respect to depth constructed for each subregion. The thresholds were detected based on changes in mean and/or variance of the partial residuals. Each subregion was post-stratified into a maximum of three depth strata separately for each year and for abundance and biomass data.

We estimated the mean abundance (or biomass) and its variance by segment and stratum using Eqs. 9 and 10. An example of the calculated mean and CVs for each segment for 2015 biomass data is in Fig. 7. These mean and variances were weighted by total field of view (f) and length of the segment (d) to estimate the total abundance or biomass and its variance

$$\hat{T} = A \sum_{i=1}^{3} S_i \sum_{j=1}^{n_i} w_{i,j} \bar{t}_{i,j} \tag{18}$$

$$\mathrm{Var}\left(\hat{T}\right) = A^2 \sum_{i=1}^{3} S_i^2 \sum_{j=1}^{n_i} w_{i,j}^2 \mathrm{Var}\left(\bar{t}_{i,j}\right), \tag{19}$$

where n_i is number of segments within depth stratum i, S_i is the size of depth stratum i, and $w_{i,j}$ is the weighting factor

$$w_{i,j} = \frac{d_{i,j} - d_{i,(1)}}{2\left(d_{i,(n_i)}\right) - d_{i,(1)}}) + \frac{f_{i,j} - f_{i,(1)}}{2\left(f_{i,(n_i)} - f_{i,(1)}\right)}, \tag{20}$$

where the numbers in parentheses represent order statistics.

Assessment

Model testing using simulations

The proportion of converged model runs was 99% for GAM + OK and OK but only 52–72% for GAMM + OK (Table 2). The type of simulated population and survey track did not affect the optimal model or segment length; results were thus not separated by these factors. Among the three model-

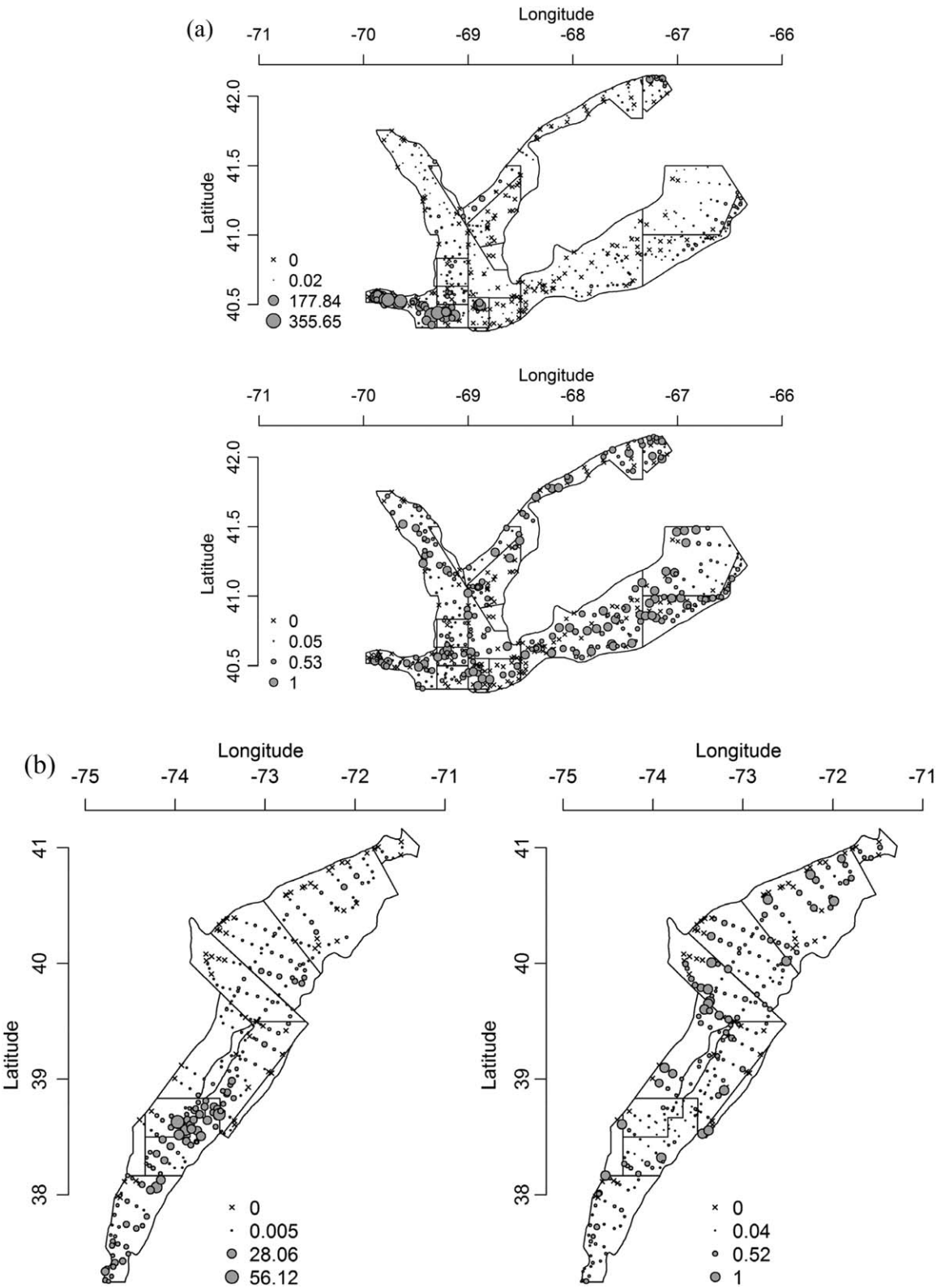

Fig. 7. Mean biomass (top or left; g/m^2) and CV (bottom or right) by segments for 2015 for (**a**) GB and (**b**) MAB.

Table 2. Summary of RBias and RRMSE for mean and CV, nugget/sill (N/S) ratio, and percent of converged sample runs (out of 7200) for biomass and abundance estimates by segment length and estimation methods.

		Biomass					Abundance				
Model type	Segment length (km)	RBias (mean)	RRMSE (mean)	RBias (CV)	N/S ratio	% converged	RBias (mean)	RRMSE (mean)	RBias (CV)	N/S ratio	% converged
GAM	0.75	0.04	0.19	−0.12	0.28	99.90	0.03	0.19	−0.11	0.22	99.94
GAMM	0.75	0.08	0.20	0.21	0.28	71.94	0.07	0.19	0.38	0.23	58.46
OK	0.75	0.14	0.25	0.60	0.18	99.93	0.14	0.24	0.58	0.12	99.97
GAM	1.5	0.05	0.21	0.12	0.25	99.79	0.04	0.20	0.15	0.19	99.78
GAMM	1.5	0.08	0.20	0.95	0.27	64.32	0.08	0.20	1.32	0.22	52.13
OK	1.5	0.15	0.26	0.49	0.12	99.47	0.15	0.25	0.45	0.09	99.43
GAM	2.25	0.06	0.22	0.46	0.21	99.75	0.06	0.21	0.88	0.17	99.67
GAMM	2.25	0.08	0.21	0.67	0.26	70.36	0.08	0.20	1.39	0.20	61.57
OK	2.25	0.16	0.28	0.29	0.10	99.63	0.16	0.27	0.27	0.07	99.43
SM		0.00	0.18	−0.61			0.00	0.17	−0.66		
SMN		0.23	0.35	−0.67			0.22	0.34	−0.72		
SMW		0.14	0.25	−0.60			0.14	0.25	−0.66		

Table 3. The frequency of model type and segment lengths that best fit observations of dredge, drop camera, or out-of-sample HabCam data for the 14 subregions in GB and MAB for 2013.

Segment length (km) and model type	HabCam	NEFSC (dredge)	SMAST (drop camera)	VIMS (dredge)	Total
0.5	2	1		5	8
0.75	3	1		3	7
1	2	3	2	1	8
1.25		5			5
1.5	2	3	2		7
1.75	5			1	6
GAM	5	8	2	6	21
GAMM	7	5	1	3	16
OK	2		1	1	4

based methods, GAM + OK produced the least biased and most precise abundance and biomass estimates, followed by GAMM + OK and then OK (Table 2). GAM + OK also produced the least biased CV estimates, followed by OK and then GAMM + OK (Table 2). The RBias and RRMSE of the abundance and biomass estimates tend to increase with increasing segment length (Table 2). The increase was the most for OK models and the least for GAMM + OK models. The RBias of the CV estimates tend to increase for GAM + OK and GAMM + OK models and decrease for OK models with increasing segment length (Table 2). Taking both bias and precision of the estimates into account, the GAM + OK with 0.75 km segments was the best performing model-based method, and also produced the least biased CV estimates for both biomass and abundance estimates.

The RBias and RRMSE for the properly stratified SM estimates were smaller than all the model-based estimates but the CVs were highly underestimated (Table 2). Additionally, SM estimates were sensitive to the quality of post-stratification; SMW and SMN estimates were more biased and less precise than most of the model-based estimates (Table 2).

Model testing using 2013 HabCam biomass data

Both model fittings and validations showed that no single modeling approach was always superior, but GAM + OK generally performed the best, followed by GAMM + OK and then OK (Table 3). The models performed slightly better when segment length is small.

Application: 2011–2015 HabCam data analysis

Based on the simulation and the 2013 field data analysis results, the GAM + OK method with 0.75 km aggregation length was used to estimate total abundance and biomass for each subregion in GB and MAB for 2011–2015. We also

Table 4. Estimates of abundance and biomass from 2011 to 2015 HabCam data using the GAM + OK and SM methods, compared to concurrent scallop dredge survey estimates for these years. The 2015 GB dredge survey estimate is not available due to lack of coverage of some high density areas.

		Number (million)						Weight (mt)					
Stock	Year	SM	GAM+OK	Dredge	SM CV	GAM+OK CV	Dredge CV	SM	GAM+OK	Dredge	SM CV	GAM+OK CV	Dredge CV
GB	2011	3992	3832	4122	0.02	0.31	0.42	110204	102819	105487	0.02	0.12	0.11
GB	2012	4003	4642	3761	0.03	0.14	0.20	94025	94040	87520	0.03	0.08	0.12
GB	2013	3562	4049	4812	0.03	0.09	0.24	54683	49671	76587	0.03	0.29	0.15
GB	2014	6199	5863	7279	0.03	0.34	0.48	75805	73495	69747	0.03	0.13	0.22
GB	2015	26797	20058		0.03	0.16		179408	144151		0.03	0.14	
MAB	2012	4166	4902	4017	0.03	0.13	0.21	50574	49196	49418	0.04	0.12	0.10
MAB	2013	5064	4611	4066	0.05	0.07	0.12	62315	61485	59456	0.04	0.13	0.10
MAB	2014	5953	6459	4020	0.03	0.10	0.17	97756	91830	64400	0.04	0.08	0.12
MAB	2015	23884	20988	9906	0.02	0.20	0.12	122206	126697	98467	0.02	0.06	0.09

provided SM estimates with careful stratifications (although the CV of the SM estimates are probably understated) as well as the stratified mean estimates from NEFSC and VIMS dredge surveys to validate the model-based estimates (Hart and Rago 2006; NEFSC 2014). GAM + OK, SM, and dredge abundance and biomass estimates and their CVs for both stocks for 2011–2015 are in Table 4, and an example of the interpolation surface for 2015 biomass is in Fig. 8. GAM + OK estimates agreed well with SM estimates, except for one subregion each in 2014 and 2015. Simple linear regressions of GAM + OK estimates against SM estimates for all years and subregions without the two outliers gave intercepts and slopes of −123.93 and 1.13 for the abundance estimates and 260.14 and 1.01 for the biomass estimates (with the two outliers the intercept and slope are −282.03 and 1.34 for the abundance estimates and −1820.47 and 1.21 for the biomass estimates). GAM + OK and SM annual estimates for both stocks are similar to the dredge estimates, except for the estimates for MAB in 2015. The calculated CVs of the SM estimates were lower than the CVs of GAM + OK for both abundance and biomass.

Discussion

Our work highlights the importance of incorporating large-scale trends in spatial distribution modeling when such trends exist. The large-scale trends that occur in both the simulated and real data violate one of the basic assumptions of OK, namely that the population is (weakly) spatially stationary. As a result, the simulations and field data analysis demonstrated that OK estimates were biased and imprecise because they do not account for the large-scale trends. RK, in our case using GAM + OK or GAMM + OK models, can account for these trends and was performed better than OK. Similar conclusions were reached in studies of fish (Yu et al. 2013), soil (Knotters et al. 1995; Odeh et al. 1995), and solar radiation distributions (Alsamamra et al. 2009). In most cases, the GAM + OK approach usually performed better than the GAMM + OK, but the reasons for this is unclear; one possibility is that the algorithms used to estimate GAMMs are not giving as stable or as reliable estimates as GAMs using only fixed effects.

The OK method was the most sensitive to the aggregation/grouping length, followed by GAM + OK and then GAMM + OK. The reduced sensitivity of GAMM models is probably because it treats small-scale variations as random effects, rather than simply averaging the data within each segment. The estimated nugget/sill ratios decreased more by segment length for GAM + OK and OK but less so for GAMM + OK models (Table 2). Averaging data could reduce the resolution of local information and smear small-scale spatial structures. This could cause misrepresentations of local conditions when the segment length is larger than the natural aggregations of target species, especially when the aggregations are relatively small and dense. The increase of bias and decrease of precision with increasing segment length is more severe for the simulated population with smaller ranges and denser patches (Table 5). This is a particular problem for OK since the large-scale trends are not modeled.

Anisotropy had relatively less influence on the biomass and abundance estimates compared to the model type and the segment length. Much of the anisotropy in the data was induced by large-scale trends. For this reason, the OK method was more sensitive to anisotropy than the RK approaches. Ignoring anisotropy can cause inaccurate and imprecise abundance and biomass estimates, especially when OK alone was used. For example, the anisotropic OK models for biomass in the northern flank area of GB in 2013 varied by as much as 40%.

Although the RK approach performed well in our analysis, it was criticized by Cressie (1993) and Lark et al. (2006)

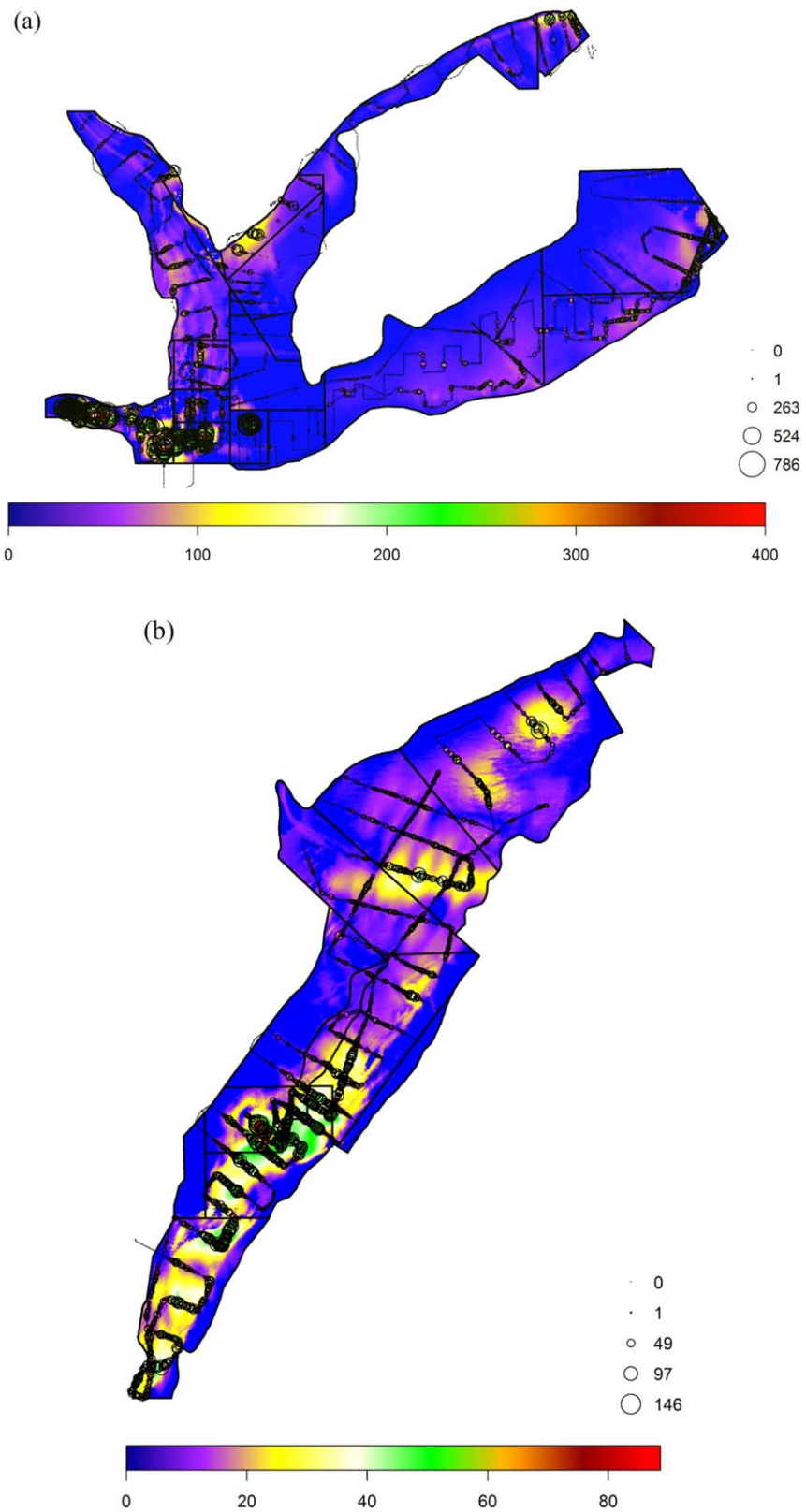

Fig. 8. Estimated scallop biomass (mt/km^2) on (**a**) GB and (**b**) MAB in 2015 based on HabCam data using GAM+ OK method. The observations of scallops (black circles; area proportional to g/m^2) are also shown.

Table 5. Summary of difference in RBias and RRMSE for mean biomass and abundance estimates by population type, relative to 0.75 km segment length.

Population type		Biomass				Abundance			
		Diff. RBias		Diff. RRMSE		Diff. RBias		Diff. RRMSE	
First-order effect	Second-order effect	1.5–0.75	2.25–0.75	1.5–0.75	2.25–0.75	1.5–0.75	2.25–0.75	1.5–0.75	2.25–0.75
Narrow	Large Agg. (noise)	0.0011	0.0076	0.0017	0.0092	0.0049	0.0119	0.0064	0.0210
Wide	Large Agg. (noise)	0.0039	0.0066	0.0056	0.0089	0.0049	0.0124	0.0035	0.0165
Narrow	Small Agg. (noise)	0.0067	0.0193	0.0085	0.0273	0.0069	0.0196	0.0131	0.0273
Wide	Small Agg. (noise)	0.0056	0.0144	0.0077	0.0185	0.0055	0.0145	0.0062	0.0144
Narrow	Large Agg.	0.0076	0.0185	0.0088	0.0262	0.0099	0.0182	0.0089	0.0212
Wide	Large Agg.	0.0104	0.0208	0.0125	0.0236	0.0093	0.0172	0.0071	0.0138
Narrow	Small Agg.	0.0177	0.0322	0.0217	0.0436	0.0234	0.0375	0.0199	0.0348
Wide	Small Agg.	0.0135	0.0260	0.0198	0.0356	0.0117	0.0184	0.0108	0.0158

because the variogram estimates of the random component of the spatial variation are theoretically biased. Generalized least squared and residual maximum likelihood-empirical best linear unbiased predictors are the two potential solutions (Lark et al. 2006). However, Kitanidis (1993) and Minasny and McBratney (2007) showed that while these methods are theoretically preferable to RK, they did not substantially improve model predictions.

The spatial autocorrelation inherent in the data can be handled in the model-based methods by incorporating the spatial structures into modeling framework using, e.g., kriging or Gaussian random fields. However, this can be a problem for design-based estimators because the samples are not randomized. Even after the samples were aggregated into stratified segments, successive segments in each stratum were still somewhat correlated. The positive correlation between successive observations causes the naïve variance estimator of the sample mean, which assumes the samples to be independent, to be biased low, with the degree of underestimation depending on the strength of the correlation (Cochran 1977). Williamson (1982) suggested using the cluster sampling variance estimators (Hansen and Hurwitz 1953), but this method is an approximation and only effective when the CV for the cluster is lower than 0.2. Because more than half of the CVs of segments of the HabCam data were larger than 0.2 for most years, we still used the naïve variance estimates for this study. Even though the tracks were carefully segmented to break down the spatial autocorrelation, the simulation results indicated that the sample variance was still underestimated.

The simulations indicated that the design-based estimator (SM) performed as well or even better than the best model-based estimates when the study area is well-stratified. However, because the higher density central portion was sampled at a higher intensity than the lower density areas, incorrect stratifications (i.e., making the central stratum too wide or too thin) will cause the stratified mean to be biased high. Seriously inaccurate stratifications that are not consistent with the survey design resulted in biased and imprecise estimates that were worse than almost all the model-based estimates. This could be avoided by surveying all areas at the same rate, but at the cost of reduced precision since more time would be spent in low density areas that contribute little to the overall mean.

In practice, SM estimates generally agreed well with the GAM + OK model-based estimates, except for two cases, where the SM estimates were 20–40% higher than those from GAM + OK. Most of the survey transects in these cases were placed in the areas with high densities of scallops with few in the marginal habitats, which made it difficult to properly stratify the area. Influences of stratification on design-based estimates have been documented in several studies. For example, Brandt et al. (1991) showed a 20% difference for horizontally stratified estimates and vertically stratified estimates for abundance estimates based on acoustic data for pelagic fish in Lake Michigan.

Our model-based and SM estimates from HabCam data generally agreed well with the estimates from the dredge data, with the exception of MAB in 2015, where HabCam estimates were substantially higher than the dredge estimates. A very large year classes of juvenile scallops were observed that year (Figs. 7 and 8). Dredge efficiency may be reduced in the presence of these high densities, thereby inducing underestimates in the dredge survey.

Our results indicate that RK using GAM models, with a relatively short aggregation length, is an accurate and precise method to estimate HabCam data or other similar data sets. The stratified mean approach is also effective provided that the (post) stratification can be done precisely, for example if accurate stratification can be built into the survey design. Using OK alone is not recommended if there are substantial large-scale trends in the data.

References

Alsamamra, H., J. A. Ruiz-Arias, D. Pozo-Vázquez, and J. Tovar-Pescador. 2009. A comparative study of ordinary and residual kriging techniques for mapping global solar radiation over southern Spain. Agric. For. Meteorol. **149**: 1343–1357. doi:10.1016/j.agrformet.2009.03.005

Barry, S. C., and A. H. Welsh. 2002. Generalized additive modeling and zero-inflated count data. Ecol. Modell. **157**: 179–188. doi:10.1016/S0304-3800(02)00194-1

Brand, A. R. 2006. Scallop ecology: Distributions and behavior, p. 651–744. *In* S. E. Shumway and G. J. Parsons [eds.], Scallops: Biology, ecology and aquaculture. Elsevier.

Brandt, S. B., D. M. Mason, E. V. Patrick, R. L. Argyle, L. Wells, P. A. Unger, and D. J. Stewart. 1991. Acoustic measures of the abundance and size of pelagic planktivores in Lake Michigan. Can. J. Fish. Aquat. Sci. **48**: 894–908. doi:10.1139/f91-106

Cochran, W. G. 1977. Sampling techniques. John Wiley and Sons.

Cressie, N. A. C. 1993. Statistics for spatial data, revised edition 68. John Wiley and Sons.

Georgakarakos, S., and D. Kitsiou. 2008. Mapping abundance distribution of small pelagic species applying hydroacoustics and Co-Kriging techniques. Hydrobiologia **612**: 155–169. doi:10.1007/s10750-008-9484-z

Hansen, M. H., and W. N. Hurwitz. 1953. Sample survey methods and theory, v. I. and II. Wiley.

Hart, D. R. 2006. Effects of sea stars and crabs on sea scallop (*Placopecten magellanicus*) recruitment in the Mid-Atlantic Bight (USA). Mar. Ecol. Prog. Ser. **306**: 209–221. doi:10.3354/meps306209

Hart, D. R., and P. J. Rago. 2006. Long-term dynamics of US Atlantic sea scallop *Placopecten magellanicus* populations. N. Am. J. Fish. Manag. **26**: 490–501. doi:10.1577/M04-116.1

Hedley, S. L., and S. T. Buckland. 2004. Spatial models for line transect sampling. J. Agric. Biol. Environ. Stat. **9**: 181–199. doi:10.1198/1085711043578

Hengl, T. 2009. A practical guide to geostatistical mapping. Univ. of Amsterdam.

Hennen, D. R., and D. R. Hart. 2012. Shell height-to-weight relationships for Atlantic sea scallops (*Placopecten magellanicus*) in offshore U.S. Waters. J. Shellfish Res. **31**: 1133–1144. doi:10.2983/035.031.0424

Howland, J., S. Gallager, H. Singh, A. Girard, L. Abrams, C. Griner, R. Taylor, and N. Vine. 2006. Development of a towed survey system for deployment by the fishing industry. IEEE Oceans 2006, 5 p.

Jolly, G. M., and I. Hampton. 1990. A stratified random transect design for acoustic surveys of fish stocks. Can. J. Fish. Aquat. Sci. **47**: 1282–1291. doi:10.1139/f90-147

Killick, R., I. A. Eckley, P. Jonathan, and K. Ewans. 2010. Detection of changes in the characteristics of oceanographic time-series using statistical change point analysis. Ocean Eng. **37**: 1120–1126. doi:10.1016/j.oceaneng.2010.04.009

Kitanidis, P. K. 1993. Generalized covariance functions in estimation. Math. Geol. **25**: 525–540. doi:10.1007/BF00890244

Knotters, M., D. J. Brus, and J. O. Voshaar. 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma **67**: 227–246. doi:10.1016/0016-7061(95)00011-C

Lark, R. M., B. R. Cullis, and S. J. Welham. 2006. On spatial prediction of soil properties in the presence of a spatial trend: The empirical best linear unbiased predictor (E-BLUP) with REML. Eur. J. Soil Sci. **57**: 787–799. doi:10.1111/j.1365-2389.2005.00768.x

Maravelias, C. D., D. G. Reid, E. J. Simmonds, and J. Haralabous. 1996. Spatial analysis and mapping of acoustic survey data in the presence of high local variability: Geostatistical application to North Sea herring (*Clupea harengus*). Can. J. Fish. Aquat. Sci. **53**: 1497–1505. doi:10.1139/f96-079

Mello, L. G., and G. A. Rose. 2005. Using geostatistics to quantify seasonal distribution and aggregation patterns of fishes: An example of Atlantic cod (*Gadus morhua*). Can. J. Fish. Aquat. Sci. **62**: 659–670. doi:10.1139/f04-227

Minasny, B., and A. B. McBratney. 2007. Spatial prediction of soil properties using EBLUP with the Matérn covariance function. Geoderma **140**: 324–336. doi:10.1016/j.geoderma.2007.04.028

Northeast Fisheries Science Center (NEFSC). 2010. 50th Northeast Regional Stock Assessment Workshop: Assessment Report. Northeast Fisheries Science Center Reference Document 10-17.

Northeast Fisheries Science Center (NEFSC). 2014. 59th Northeast Regional Stock Assessment Workshop: Assessment Report. Northeast Fisheries Science Center Reference Document 14-09.

Odeh, I. O. A., A. B. McBratney, and D. J. Chittleborough. 1995. Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression-kriging. Geoderma **67**: 215–226. doi:10.1016/0016-7061(95)00007-B

Páramo, J., and R. Roa. 2003. Acoustic-geostatistical assessment and habitat–abundance relations of small pelagic fish from the Colombian Caribbean. Fish. Res. **60**: 309–319. doi:10.1016/S0165-7836(02)00142-X

Petitgas, P. 1993. Geostatistics for fish stock assessments: A review and an acoustic application. ICES J. Mar. Sci. **50**: 285–298. doi:10.1006/jmsc.1993.1031

Rivoirard, J., J. Simmonds, K. G. Foote, P. Fernandes, and N. Bez. 2008. Geostatistics for estimating fish abundance. John Wiley & Sons.

Särndal, C. E., I. Thomsen, J. M. Hoem, D. V. Lindley, O. Barndorff-Nielsen, and T. Dalenius. 1978. Design-based and model-based inference in survey sampling. Scand. J. Stat. **5**: 27–52.

Simard, Y., D. Marcotte, and G. Bourgault. 1993. Exploration of geostatistical methods for mapping and estimating acoustic biomass of pelagic fish in the Gulf of St. Lawrence: Size of echo-integration unit and auxiliary environmental variables. Aquat. Living Resour. **6**: 185–199. doi:10.1051/alr:1993020

Simard, Y., and D. Lavoie. 1999. The rich krill aggregation of the Saguenay-St. Lawrence Marine Park: Hydroacoustic and geostatistical biomass estimates, structure, variability, and significance for whales. Can. J. Fish. Aquat. Sci. **56**: 1182–1197. doi:10.1139/cjfas-56-7-1182

Singh, W., E. B. Örnólfsdóttir, and G. Stefansson. 2013. A camera-based autonomous underwater vehicle sampling approach to quantify scallop abundance. J. Shellfish Res. **32**: 725–732. doi:10.2983/035.032.0314

Smith, A. N., M. J. Anderson, and R. B. Millar. 2012. Incorporating the intraspecific occupancy-abundance relationship into zero-inflated models. Ecology **93**: 2526–2532. doi:10.1890/12-0460.1

Smith, S. J. 1990. Use of statistical models for the estimation of abundance from groundfish trawl survey data. Can. J. Fish. Aquat. Sci. **47**: 894–903. doi:10.1139/f90-103

Stokesbury, K. D., B. P. Harris, M. C. Marino, and J. I. Nogueira. 2004. Estimation of sea scallop abundance using a video survey in off-shore US waters. J. Shellfish Res. **23**: 33–41.

Taylor, R., and others. 2008. Evolution of a benthic imaging system from a towed camera to an automated habitat characterization system. IEEE Oceans 2008, 7 p.

Webster, R., and M. A. Oliver. 2001. Geostatistics for environmental scientist. John Wiley and Sons.

Wedderburn, R. W. M. 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton Method. Biometrika **61**: 439–447. doi:10.1093/biomet/61.3.439

Williams, R. O. B., S. L. Hedley, T. A. Branch, M. V. Bravington, A. N. Zerbini, and K. P. Findlay. 2011. Chilean blue whales as a case study to illustrate methods to estimate abundance and evaluate conservation status of rare species. Conserv. Biol. **25**: 526–535. doi:10.1111/j.1523-1739.2011.01656.x

Williamson, N. J. 1982. Cluster sampling estimation of the variance of abundance estimates derived from quantitative echo sounder surveys. Can. J. Fish. Aquat. Sci. **39**: 229–231. doi:10.1139/f82-031

Yu, H., Y. Jiao, and L. W. Carstensen. 2013. Performance comparison between spatial interpolation and GLM/GAM in estimating relative abundance indices through a simulation study. Fish. Res. **147**: 186–195. doi:10.1016/j.fishres.2013.06.002

Zuur, A. F., A. A. Saveliev, and E. N. Ieno. 2012. Zero-inflated models and generalized linear mixed models with R. Highland Statistics Ltd.

Acknowledgments

We thank the NEFSC Ecosystems Survey Branch, S. Gallager, A. York, R. Taylor, K. Boles, and N. Vine for collection of the HabCam data, D. Rudders and K. Stokesbury for providing dredge and drop camera survey data, and D. Hennen, M. Simpkins and an anonymous reviewer for their constructive comments on the manuscript.

Conflict of Interest

None declared.

Submitted 24 March 2016
Revised 10 November 2016
Accepted 26 January 2017

Associate editor: Craig Lee