Supplementary materials for this article are available at 10.1007/s13253-017-0295-4.

Bayesian Calibration of Blue Crab (*Callinectes sapidus*) Abundance Indices Based on Probability Surveys

Dong LIANG, Genevieve NESSLAGE, Michael WILBERG, and Thomas MILLER

Abundance and standard error estimates in surveys of fishery resources typically employ classical design-based approaches, ignoring the influences of non-design factors such as varying catchability. We developed a Bayesian approach for estimating abundance and associated errors in a fishery survey by incorporating sampling and non-sampling variabilities. First, a zero-inflated spatial model was used to quantify variance components due to non-sampling factors; second, the model was used to calibrate the estimated abundance index and its variance using pseudo empirical likelihood. The approach was applied to a winter dredge survey conducted to estimate the abundance of blue crabs (*Callinectes sapidus*) in the Chesapeake Bay. We explored the properties of the calibration estimators through a limited simulation study. The variance estimator calibrated on posterior sample performed well, and the mean estimator had comparable performance to design-based approach with slightly higher bias and lower (about 15% reduction) mean squared error. The results suggest that application of this approach can improve estimation of abundance indices using data from design-based fishery surveys.

Key Words: Auxiliary information; Empirical likelihood; Integrated Nested Laplace Approximation (INLA); Model-assisted approach; Survey design; Index standardization; Variance estimation.

1. INTRODUCTION

Fisheries managers commonly make resource decisions by comparing estimates of current population abundance to reference levels. Consequently, the accuracy of estimated abundance is an important consideration for managers in interpreting changes in population size (Chen et al. 2004; Wagner et al. 2007). Scientific surveys that collect data independently of the fishery itself, termed fishery-independent surveys, are the preferred basis for management. Kimura and Somerton (2006) provide an introduction to the large statistical literature

Dong Liang Environmental Statistics Collaborative, Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, 146 Williams Street, Solomons, MD 20688, USA. Genevieve Nesslage, Michael Wilberg and Thomas Miller (✉) Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, 146 Williams Street, Solomons, MD 20688, USA (E-mail: *miller@umces.edu*).

Journal of Agricultural, Biological, and Environmental Statistics, Volume 22, Number 4, Pages 481–497
DOI: 10.1007/s13253-017-0295-4

on fishery surveys that has been developed. On occasion estimates of abundance derived directly from data collected during the commercial or recreational fishery, and thus termed fishery-dependent, must be used. However, such estimates often perform poorly (Maunder and Punt 2004). In both fishery-independent and fishery-dependent approaches, the catch per unit of effort (CPUE) in the survey serves as an index of abundance (Chen et al. 2004; Kimura and Somerton 2006).

Fishery-independent surveys rely on an underlying design, often a stratified random sample (Kimura and Somerton 2006). Two approaches have been used to estimate abundance from the survey data: design-based and model-based estimates, sometimes with standardization for covariates (Särndal et al. 1978; Smith 1990). The classical design-based approach has several advantages in deriving unbiased estimates, especially for routine analyses of many variables (Opsomer et al. 2007). Although design-based estimates of the fishery-independent surveys may represent the "best scientific information available" regarding population status, the design-based standard error only accounts for the uncertainty due to probability-based sampling design. Even with rigorous survey designs and the use of standardized gear in fishery-independent surveys, non-design factors may have substantial impacts on estimates. One example of a non-design factor known to have a substantial impact on abundance estimates is survey catchability, the proportionality constant between an index of abundance and population size (Kimura and Somerton 2006; Wilberg et al. 2010) .

When non-sampling factors are important, a model-assisted framework allows estimation of variance components due to non-sampling factors such as the catching process, sightability (Fieberg et al. 2013), and environmental conditions (Valliant et al. 2000; Chen et al. 2004). There are many potential statistical frameworks for model-assisted approaches. The pseudo-empirical likelihood (EL) approach has received substantial interest in analyzing survey data (Chen and Sitter 1999; Chen et al. 2004). The EL estimator is model unbiased under informative survey designs (Pfeffermann et al. 1998; Pfeffermann 2007; Savitsky and Toth 2016) and its robustness to model misspecification under mild regularity conditions makes it naturally suitable for incorporating auxiliary information (Wu and Sitter 2001). For spatially intensive surveys, the geographic information collected as a part of the survey may provide valuable information regarding the species' distribution within the sampling domain (Jensen and Miller 2005 and references therein). The contribution of spatial information to EL-based abundance estimation has not been explored (Brus and DeGruijter 1993).

Bayesian approaches using Markov chain Monte Carlo (MCMC) samples are well-positioned to estimate variance components due to non-sampling factors. The stochastic partial differential equation approach (Lindgren et al. 2011) can be used to incorporate geospatial information in the inferential framework when combined with the integrated nested Laplace approximation (INLA, Rue et al. 2009) to generate approximate samples from a posterior distribution. The INLA approach is particularly attractive for larger spatial data analyses where conventional MCMC algorithms require computationally intensive matrix operations (Banerjee et al. 2014). Using INLA, the total variance in the data can be used in the estimation of the variance components.

Our objectives were to develop a modeling framework to estimate an index of abundance and associated variance estimates from CPUE data collected in a fishery survey. We evalu-

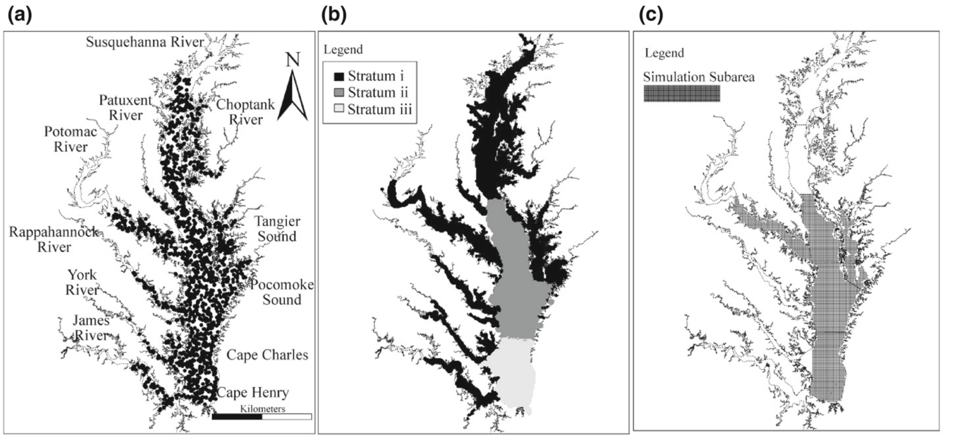

Figure 1. Winter dredge survey area, Chesapeake Bay, United States: **a** A typical distribution of sample locations during winter 1997/1998. **b** Strata design used in the survey since 1994. Stratum 1- Upper Bay and rivers, Stratum 2- Middle Bay, Stratum 3 - Lower Virginia Bay. **c** Subarea used in simulation study.

ated a semi-continuous, or zero-inflated (Liu et al. 2011; Thorson and Ward 2013) spatial model that relates expected CPUE to auxiliary information based on design variables and geographic locations. Also, we conducted a limited simulation study to investigate the performance of the proposed estimators, using a synthetic population based on the survey data. We applied our techniques to survey data for an exploited marine crustacean, the blue crab (*Callinectes sapidus*) in the Chesapeake Bay, U.S. (Sharov et al. 2003; Fig. 1a). Although we used a blue crab survey in this example, our approach is readily applicable for estimating abundances and associated variances for other design-based and spatially explicit surveys of aquatic and marine populations.

2. MATERIALS AND METHODS

In Sect. 1, we define a crab abundance index based on CPUE. In Sects. 2 and 3, respectively, we introduce a motivation data set and propose a working model for CPUE. In Sects. 4 and 5, respectively, we implement a Bayesian EL estimator for an index of abundance and that for the associated variance estimator.

2.1. DEFINITION OF CRAB ABUNDANCE INDEX

Populations of exploited marine resources exhibit frequent and often substantial changes in abundance due to migration, recruitment, and both natural and fishing mortality. These processes challenge the conventional survey sampling framework, which assumes a closed population and a static catch process (i.e., constant catchability). Thus, the estimated CPUE reflects both the population status as well as the sampling or the catching process and environmental conditions (Wilberg et al. 2010).

Given standard gear and a short time frame, we may assume CPUE is a smooth function of the covariates **x** (including spatial location), which facilitates a spatially explicit modeling approach (Jensen and Miller 2005). Following the notation of Chen et al. (2004), let R denote the study region and $\mu(\mathbf{x})$ the expected CPUE for location $\mathbf{x}$. Theoretically, we can define the abundance index as $I(R) = \int_R \mu(\mathbf{x})\mathrm{d}\mathbf{x}$. For computation ease, if we let g_i, $i = 1, \ldots, N$, denote a tessellation of size N in domain R, we can define

$$I(R) = \sum\nolimits_{i \leq N} \mu(\mathbf{x}_i). \tag{1}$$

In practice, we implement a sampling design p and observe Y_i, the variable of interest which in this case is CPUE, within n finite grids, assuming standard gear usage and regular environmental conditions. Thus, we assume that

$$\mathrm{E}_\xi\left(Y_i|\mathbf{x}_i\right) = \mu(\mathbf{x}_i), \tag{2}$$

where ξ denotes the latent stochastic model that generates the observed data; ξ quantifies the uncertainty due to the varying catchability. We wish to derive an estimator $\hat{I}(R)$ that is an unbiased estimator of $I(R)$, i.e., $\mathrm{E}_\xi\,\mathrm{E}_p[\hat{I}(R)] = I(R)$ (Smith 1990).

2.2. Winter Dredge Survey

A stratified random winter dredge survey (WDS) has been conducted for blue crab in the Chesapeake Bay between December and March since the winter of 1989/1990 (Sharov et al. 2003, henceforth, we refer to each survey by the year in which it was completed). Between 900 and 1,500 stations (Fig. 1a) in waters deeper than 1.5 meters have been sampled each year. Winter was selected for the survey because crabs are buried in the sediment and largely dormant during this time of the year, and thus exhibit limited movement. This limited movement and the short duration of the survey mean that the population can be considered closed.

The sampling design has been consistent since 1994. Three fixed geographic strata have been employed based on salinity (Sharov et al. 2003): (i) the upper Bay and rivers; (ii) the middle Bay and (iii) the lower Bay (Fig 1b). The number of stations in each geographic stratum has been proportional to stratum area. Thus, from a spatial viewpoint, the design is equivalent to simple random sampling with replacement (SRSWr).

The response variable Y_i, the CPUE of blue crab, was calculated as the number of blue crab (>15 mm carapace width) divided by the area swept for each tow. At each station, a 1.83-meter wide dredge was towed along the bottom at a speed of 3 knots, retaining crabs greater than 15 mm in carapace width. The geographic coordinates at the beginning and end of the dredge, projected to the UTM (zone 18N), were used to calculate the tow distance and the area swept. We included the projected latitude and bottom depth at the beginning of the dredge as numerical covariates, and stratum as a categorical covariate in the model proposed in Sect. 2.3. The sampling date was not included as a covariate because for logistical efficiency the locations were sampled sequentially; hence, the fixed effect of sampling sequence is confounded with the spatial random effects. To minimize uncertainty in tow length, only tows with dredge distance between 50 and 500 m were included in the analyses (Jensen and Miller 2005).

For the purpose of predicting the expected CPUE, the entire survey area ($9,812\,\mathrm{km}^2$) was rasterized into 20,770 cells, with coarser resolution of 1 km in the main stem of the Bay, 250 m resolution in the mesohaline estuary, and <100 m resolution in oligohaline creeks. To calculate inclusion probability, we assumed the grid points were sampled via SRSWr. The covariates (depth and stratum) are readily available for predictive inference from the baywide bathymetry data and the definition of management segments over the Chesapeake Bay (Chesapeake Bay Program 2016).

2.3. WORKING MODEL FOR CPUE

A two-stage, overdispersed, and zero-inflated model is commonly used to model fishery survey data given the often large number of samples that catch no individuals of the target species (Jensen and Miller 2005; Maunder and Punt 2004). CPUE Y_i was assumed to arise from a mixture of zero and positive catches. The chance of a nonzero catch was modeled using the Bernoulli likelihood. Assuming a nonzero catch, CPUE was modeled by a Poisson likelihood truncated to positive values. Let η_1 and η_2 represent large scale effects captured by covariates. The covariates used here were $\mathbf{x}_i$= (latitude at start of dredge i, bottom depth at start of dredge i, categorical stratum indicator at dredge i). $b(\mathbf{s}_i)$ represents the spatial auto-correlation, where $\mathbf{s}_i$ denotes the geographic coordinates associated with i^{th} data point. Standard link functions were used to model the probability and expected positive CPUE

$$\operatorname{logit}\left[\mathrm{P}\left(Y_i > 0\right)\right] = \eta_1\left(\mathbf{x}_i\right) + b\left(\mathbf{s}_i\right) + \epsilon_i \tag{3}$$

$$\log\left[\mu^*\left(\mathrm{x}_i\right)\right] = \eta_2\left(\mathbf{x}_i\right) + \beta_1 b\left(\mathbf{s}_i\right) + \beta_2 \epsilon_i. \tag{4}$$

We assumed that zero inflation resulted from spatiotemporal aggregation of blue crabs, and hence applied the constraint of a monotonic relation between the linear predictors in (3) and (4) (Liu et al. 2011; Thorson and Ward 2013). The ϵ_i error term represents overdispersion through an unstructured Gaussian field. Both fields were centered on zero to allow estimation of the fixed effects, η_1 and η_2. The difference in the logit and log scales was captured by the β parameters.

The Gaussian field $\mathbf{b}$ was assumed stationary with an exponential covariance function, partial sill σ^2, and practical range h. The covariance function, based on a Gaussian random field, was approximated by the solution to the stochastic partial differential equation (Lindgren et al. 2011). This approach enables computationally efficient modeling of large data (Lindgren and Rue 2015). The unstructured Gaussian field was assumed with variance τ^2, which is similar to the nugget effect in classical geostatistical terminology (Cressie 1993). The hyper-prior for the practical range h followed a lognormal distribution with a mean equal to the natural logarithm of 20% of the study domain size and a precision of 0.1. The square root of the partial sill σ followed a lognormal prior with mean 0 and precision 0.1, which is a vague prior that covers most of the range of the spatial random effects with 95 % prior probability. The nugget variance τ^2 followed an inverse gamma (0.01, 0.01) distribution. The parameters τ^2, h, and σ were *a priori* independent. We used the R-INLA package (Lindgren and Rue 2015) for posterior inferences by generating approximate MCMC samples based on INLA (Rue et al. 2009). Convergence assessment was not performed due to the approximate nature of the algorithm and the independent sampling from the approximation.

Because the design was non-informative and self-weighting, we estimated the Bayesian spatial model without using design weights.

2.4. Model-Assisted Estimation Using INLA Model

Let y_i denote the observed CPUE in the dredge at station i, $\pi_i = \Pr(i \in s)$ denote the first order inclusion probability of grid i, and s denote the sample (Horvitz and Thompson 1952). Let $d_i = 1/\pi_i$, then $\hat{I}_{\mathrm{HT}}(R) = \sum_{i \in s} d_i y_i$ is the classical Horvitz–Thompson estimator of abundance index. The standard errors for $\hat{I}_{\mathrm{HT}}(R)$ were calculated using the usual large sample approximation. The model calibration approach was proposed to make efficient use of complete auxiliary information related to the response (Wu and Sitter 2001). The INLA model was used as an approximation to ξ (Eq. 2) in the estimation stage of the fish abundance index. Given the nonlinearity of the working model, the auxiliary information should be incorporated through the fitted values denoted by $u_i = \hat{\mu}(\mathbf{x}_i), i = 1, \ldots, N$ (Wu and Sitter 2001; Opsomer et al. 2007). The resulting model calibration estimator is $\sum_{i \in s} w_i y_i$ where the calibration weights w_i minimized the Chi-square distance (Wu and Sitter 2001) between w_i and $d_i, i = 1, \ldots, N$ subject to

$$N^{-1} \sum_{i \in s} w_i = 1, \text{ and} \sum_{i \in s} w_i u_i = \sum_{j=1}^{N} u_j. \tag{5}$$

The calibrated estimator is asymptotically design unbiased under mild regularity conditions (Wu and Sitter 2001) including a continuous and bounded function of $\mu(\mathbf{x})$.

The resulting weights, however, can be negative and can theoretically generate a negative estimate of abundance. The EL method (Chen and Sitter 1999; Chen et al. 2002, 2004) is asymptotically similar to model calibration (Wu and Sitter 2001), but guarantees positive weights. The EL method also has a clear maximum likelihood interpretation (Chen et al. 2004). In the EL framework, let $l(\mathbf{p})$ denote the pseudo-empirical likelihood function where

$$l(\mathbf{p}) = \sum_{i \in s} d_i \log(p_i). \tag{6}$$

Let $\hat{p}_i$ denote the maximizer of $l(\mathbf{p})$ subject to the constraints

$$\sum_{i \in s} p_i = 1\ (p_i > 0), \text{ and} \sum_{i \in s} p_i u_i = \frac{1}{N} \sum_{j=1}^{N} u_j. \tag{7}$$

The calibration weights are $w_i = N\hat{p}_i$. The EL estimator has the same analytical design variance formula (Wu and Sitter 2001). Hence, it can be viewed algebraically as a regression estimator using u_i as the auxiliary variable (Opsomer et al. 2007). Let $w_i^* = d_i / \sum_{i \in s} d_i$,the Lagrange multiplier method can be used to show that $\hat{p}_i = w_i^* / (1 + \lambda u_i)$ for $i = 1, \ldots, n$, and the scalar Lagrange multiplier λ is the solution to

$$\sum_{i \in s} \frac{w_i^* u_i}{1 + \lambda u_i} = 0. \tag{8}$$

The solution can be obtained through an efficient bisection search algorithm (Wu 2005).

The posterior predictive distribution is highly skewed: hence we used the median instead of mean of the posterior predictive distribution for $\hat{y}_i$, denoted by m_i, as the auxiliary variable in a linear regression model

$$y_i = m_i + \text{error}. \tag{9}$$

Let u_i denote the predicted value from equations (3) and (4), which was used in the EL framework to derive an estimate of $I(R)$. We denote the calibrated estimator

$$\hat{I}_{\text{MA1}}(R) = \sum\nolimits_{i \in s} \hat{p}_i y_i. \tag{10}$$

The estimator is asymptotically design unbiased, but not necessarily unbiased in the model sense due to the zero-inflated and overdispersed distribution of the catch data.

We considered a Bayesian model calibration estimator that was approximately model unbiased as well. Let $\left\{u_i^{(m)} : i = 1 \ldots N, m = 1, \ldots, M\right\}$ denote an approximate posterior predictive sample from the working model where M is the Monte Carlo sample size. We estimate $E_\xi \left\{E_p\left[I(R)\right]\right\}$ through Monte Carlo integration:

$$\hat{I}_{\text{MA2}}(R) = \frac{1}{M} \sum\nolimits_{m=1}^{M} \hat{I}_{\text{MA1}}^{(m)}(R). \tag{11}$$

where $\hat{I}_{\text{MA1}}^{(m)}(R)$ denotes the calibrated estimate of $I(R)$ based on m^{th} Monte Carlo sample from the posterior predictive distribution. Thus the full posterior distribution was used instead of a point estimate in the calibration process.

2.5. Variance Components Estimation Using INLA Model

The model-assisted approach can incorporate auxiliary information to improve efficiency of an estimator and can also consider multiple sources of uncertainties in estimating the variance of the estimator $\hat{I}_{\text{MA2}}(R)$, denoted by $\hat{I}(R)$ for brevity of notation in this section. Let Var denote the total variance of $\hat{I}(R)$ incorporating both the model ξ and the design p:

$$\begin{aligned} \text{Var}\left[\hat{I}(R)\right] &= \text{Var}_\xi \left\{E_p\left[\hat{I}(R)\right]\right\} + \text{E}_\xi\left\{\text{Var}_p\left[\hat{I}(R)\right]\right\} \\ &= \text{Var}_\xi\left\{\sum\nolimits_{i \le N} Y_i\right\} + \text{E}_\xi\left\{\text{Var}_p\left[\hat{I}(R)\right]\right\}. \end{aligned} \tag{12}$$

The posterior samples from the working model allow estimation of Var_ξ and E_ξ based on approximate Monte Carlo integration. Given posterior draws, the variance component due to model: $\text{Var}_\xi(\sum_{i \le N} Y_i)$ can be estimated from the sample variance of $\{\sum_{i \le N} \hat{y}_j^{(m)}$, $m = 1 \ldots, M\}$, where $\hat{y}_j^{(m)}$ denotes the predicted value from (3), (4) & (9) based on mth posterior predictive sample. Letting $\hat{I}^{(m)}(R)$ denote the model-assisted estimates based on posterior draw m, the estimate of sampling variance can be derived as (pp 294, Särndal et al. 1992)

$$\begin{aligned} \widehat{\text{Var}}_p\left[\hat{I}^{(m)}(R)\right] &= \sum\nolimits_{j=1}^{n} \left[1 - \pi\left(\mathbf{x_j}\right)\right] \frac{e_m^2\left(\mathbf{x}_j\right)}{\pi^2\left(\mathbf{x}_j\right)} \\ &\quad + \sum\nolimits_{j=1}^{n} \sum\nolimits_{j' \ne j}^{n} \frac{\pi\left(\mathbf{x}_j, \mathbf{x}_{j'}\right) - \pi(\mathbf{x}_j)\pi\left(\mathbf{x}_{j'}\right)}{\pi(\mathbf{x_j}, \mathbf{x}_{j'})} \frac{e_m(\mathbf{x}_{j'})}{\pi(\mathbf{x}_j)} \frac{e_m(\mathbf{x}_{j'})}{\pi(\mathbf{x}_{j'})} \end{aligned} \tag{13}$$

where $e_m(\mathbf{x}_j) = y_j - \hat{y}_j^{(m)}$, $\pi(\mathbf{x}_j)$ and $\pi(\mathbf{x}_j, \mathbf{x}_{j'})$ denote the first and second order inclusion probabilities. In our application section, the inclusion probabilities were calculated assuming Simple Random Sampling with Replacement (SRSWr, Thompson 2002 pp 71). The variance component due to the design can be estimated as the average of $\widehat{\mathrm{Var}}_p\left[\hat{I}^{(m)}(R)\right]$. The total variance can thus be estimated as the sum of the model and design variance components.

3. RESULTS

The estimates of baywide blue crab CPUE (>15 mm carapace width), i.e., $\hat{I}_{\mathrm{MA1}}(R)$ derived from model calibration with the median of the posterior predictive distribution in Eq.(10), were larger than stratified design-based estimates $\hat{I}_{\mathrm{HT}}(R)$ for most years except 1996, 1997 and 1999 (Table 1). In contrast, the estimates $\hat{I}_{\mathrm{MA2}}(R)$, calibrated to posterior samples, Eq. (11), were generally similar to those from design-based estimates. The standard errors (SE) due to sampling for $\hat{I}_{\mathrm{MA2}}(R)$, Eq. (13), were smaller than the corresponding design-based estimates (Table 1). In contrast, the SEs due to the stochastic process approximating the blue crab dredging process, Eq. (12), were much larger than those due to sampling (Table 1). The variance component associated with the stochastic process dominated the total variance estimates based on the Bayesian calibration. Years with high abundance index values also had high variance because of the stochastic variance component (Fig. 2). For example, the 2012 abundance had the highest point estimate, but when the total SE was considered it was not significantly different from the indices during 2008–2011 and 2013–2014 (Fig. 2).

The design-based estimates of coefficient of variation (CV) were smaller than those estimates from the Bayesian model calibration (Fig. 3). The year-to-year variation of design-based CV was also smaller than that from the model-assisted approach ($\hat{I}_{\mathrm{MA2}}$). No linear trend over time was present for design-based CV estimates (p-Value = 0.90), but a weak, positive linear trend was apparent for model-assisted CV estimates (p-Value = 0.04).

The spatial patterns of blue crab distribution showed inter-annual variability, but hot spots of CPUE density were largely consistent across years (Figs. 4 and 5; see supplemental materials for additional plots). Prediction uncertainties were high in areas of sparse samples (Fig. 4b, 5b), such as the deeper channels of the main stem of the Bay and lower Potomac River. High densities occurred in many lower Bay tributaries and eastern shore embayments (Fig. 4a, 5a).

4. SIMULATION STUDY

4.1. STUDY DESIGN

We performed a simulation study to evaluate the properties of the model calibration procedure for estimating the population mean and total variance. We focused on the lower part of the Chesapeake Bay and lower Potomac River instead of the whole Bay (Fig. 1c) to reduce the computational burden of repeatedly fitting the INLA model over a large study domain. The subarea selected represents over 60% of the sampling stations within only 15%

Table 1. Summary of Bay-wide Catch Per Unit Effort estimatesfor blue crabs (> 15 mm carapace width) in the Chesapeake Bay, U.S.

Year	N	HT		MA_1		MA_2			
		Estimate	SE	Estimate	SE	Estimate	SE_1	SE_2	SE
1994	1,413	121	7.9	155	4.9	110	22.0	5.7	22.7
1995	1,576	87	6.1	99	3.6	78	14.2	4.4	14.8
1996	1,631	196	12.8	193	7.2	169	29.0	8.7	30.3
1997	1,597	171	8.5	162	4.8	147	22.0	6.0	22.8
1998	1,592	89	5.0	121	2.8	85	9.7	3.7	10.4
1999	1,604	69	5.0	68	2.9	63	7.2	3.4	7.9
2000	1,558	61	3.7	95	2.2	64	12.7	2.8	13.0
2001	1,590	55	4.3	66	2.1	50	9.7	2.6	10.0
2002	1,581	59	4.6	113	2.3	63	8.1	2.9	8.6
2003	1,512	100	9.7	191	5.3	109	20.4	6.2	21.3
2004	1,527	75	5.5	117	3.7	71	13.9	4.1	14.5
2005	1,560	129	6.0	151	3.2	118	21.5	4.2	21.9
2006	1,537	111	5.9	219	2.7	123	28.7	3.8	29.0
2007	1,518	82	5.3	109	3.1	78	14.8	3.8	15.3
2008	1,434	88	5.7	219	2.9	109	21.1	3.9	21.4
2009	1,536	103	6.8	186	5.0	117	18.9	6.1	19.8
2010	1,521	184	8.1	267	4.4	170	26.3	5.8	26.9
2011	1,555	157	8.2	250	3.9	159	24.0	5.6	24.6
2012	1,558	250	13.1	451	6.4	274	77.7	8.9	78.2
2013	1,541	87	8.2	163	4.0	103	21.7	5.1	22.3
2014	1,545	97	6.7	147	3.4	104	28.0	4.3	28.3

HT stratified random design based estimates; MA_1 INLA model calibrated estimates $\hat{I}_{\mathrm{MA1}}$ based on the pseudo-empirical likelihood; MA_2 INLA model calibrated estimate $\hat{I}_{\mathrm{MA2}}$ obtained from Markov chain Monte Carlo samples and pseudo-empirical likelihood; SE_1 denotes standard error due to model; SE_2 denotes standard error due to design.

of the grid points. The relatively higher sampling density enabled stable estimation of the underlying CPUE density. The population index $I\ (R)$ was simulated for years 2001, 2006 and 2012, representing the years with low, medium and high abundance of blue crabs in the Chesapeake Bay, respectively.

Let N_1 denote the number of grid points in the subarea used in the simulation. Finite population values y_i, $i = 1, \ldots N_1$, were generated from the posterior predictive distribution of the fitted INLA model. Let n denote the number of stations sampled in the dredge survey in the simulation study domain during the specific year; n grid points were then drawn by simple random sampling. For each year the process of generating the finite population values and selecting the sample (three samples per population) was repeated 240 times for a total of 720 simulated data sets per year. For each sample, the standard Horvitz–Thompson estimator, the two model-assisted estimators, and their variance estimators were computed.

The calibration procedure was applied assuming the following data generating models for each sample: (1) a zero-augmented Poisson model with only fixed effects (ZAP); (2) a zero-augmented negative Binomial model with only fixed effects (ZANB); (3) a ZANB with stochastic partial differential equation (SPDE) random effects (ZANB+S); and (4) the zero-augmented Poisson model with both unstructured and SPDE random effects (ZAP+SOD), which was the "true" data generating model. Only bottom depth and latitude were considered

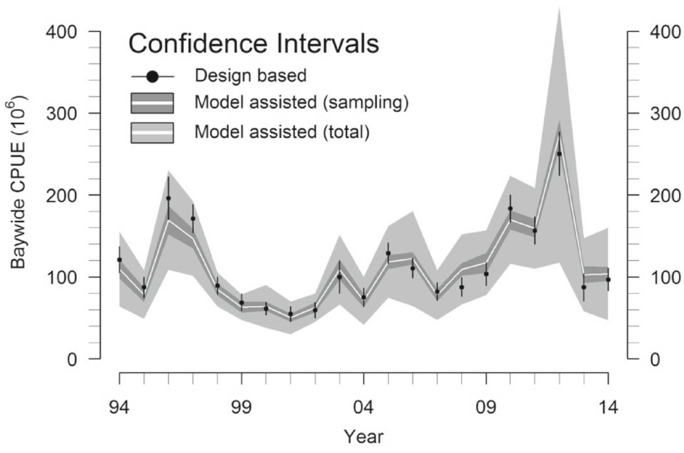

Figure 2. Time series plot of design and Bayesian calibrated estimates and standard errors (based on sampling and based on total variance including crab distribution) of Baywide CPUE (10^6).

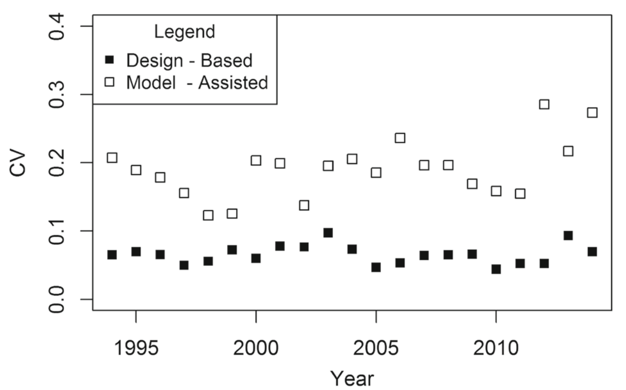

Figure 3. Time series plots of design-based and Bayesian calibrated estimates of coefficient of variation (CV) of blue crab CPUE in Chesapeake Bay, U.S.

as covariates in all four models as most of the study domain is in the lower Bay stratum (Stratum 3, Fig. 1a). For the ZANB+S model, we also chose from three hyper-priors for the sill (prior mean 0.5, 1.0 and 2.5) and three hyper-priors for the range (prior precision 0.1, 1.0 and 10), and reported results of ZANB+S from these nine combinations of hyper-priors.

The posterior mean from the fitted INLA model was used as the true CPUE density, from which we calculated the index $I(R) = N_1^{-1} \sum_{i \leq N_1} \mu(\mathbf{x}_i)$. We denote $I(R)$ as I for brevity. Let $B = 240 \times 3 = 720$ denote the total number of Monte Carlo replicates. The simulated bias and Mean Squared Error (MSE) for the estimate $\hat{I}$ were calculated as $\text{Bias}\left(\hat{I}\right) = B^{-1} \sum_{b \leq B} \left(\hat{I}_b - I\right)$ and $\text{MSE}\left(\hat{I}\right) = B^{-1} \sum_{b \leq B} \left(\hat{I}_b - I\right)^2$, respectively, where $\hat{I}_\mathrm{b}$ denotes the estimator $\hat{I}$ from the bth simulated sample. The model variance component was calculated by $V_1 = (BN_1)^{-1} \sum_{b \leq B} \sum_{i \leq N_1} (Y_{bi} - I)^2$ where Y_{bi} was the response variable from the ith grid point and the b^{th} simulation run. The variance component due to sampling was computed as $V_2 = V\text{-}V_1$ where V denote the total variance component,

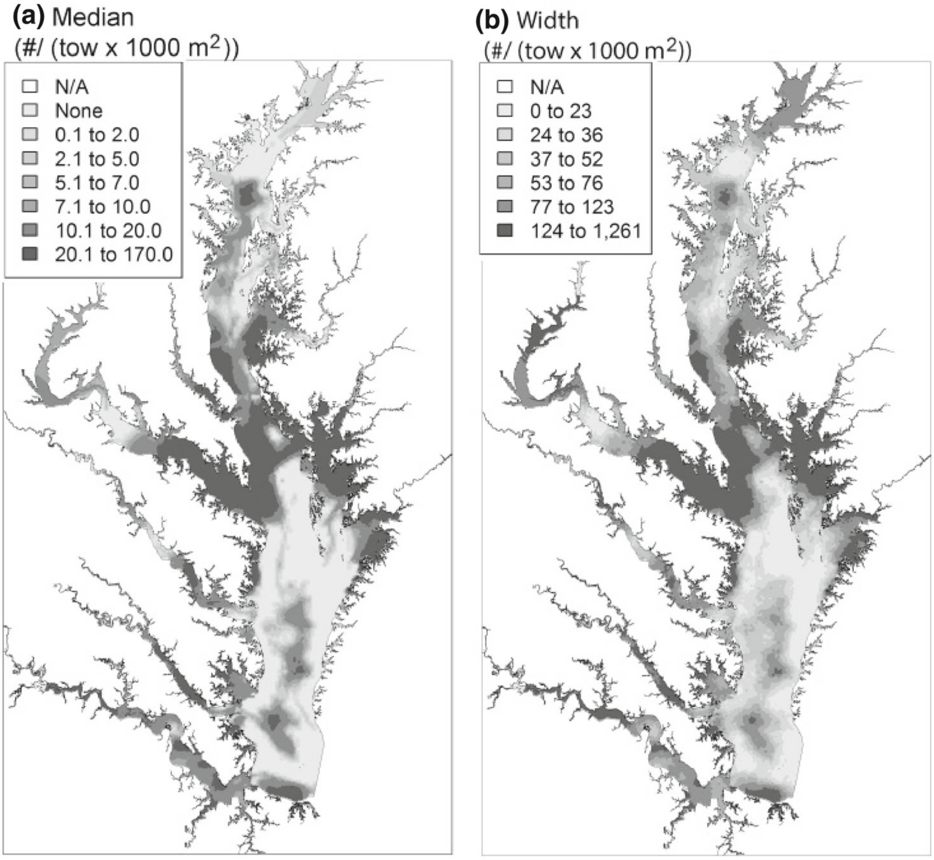

Figure 4. Maps of Chesapeake Bay during a high abundance year showing **a** blue crab winter CPUE density (number/tow/1000 m^2) and **b** inter quartile range based on spatial modeling of winter dredge survey data for winter of 2011/2012.

$V = \mathrm{MSE} - \mathrm{bias}^2$. We used $\hat{V}_1$, $\hat{V}_2$ and $\hat{V}$ to denote the average of the estimated model, design and total variance components from each simulated sample.

4.2. Simulation Results

The bias of the stratified design-based estimate was small, within 0.3% of the simulated population index (see Table 2 Column 1). The model calibration estimates based on misspecified models (ZAP & ZANB) were unbiased as well. Overall the model calibration estimators based on spatial and overdispersed models (ZAP+SOD & ZANB+S) were slightly biased. Specifically, $\hat{I}_{\mathrm{MA1}}(R)$ based on a point estimate (median) from ZAP+SOD or ZANB+S generated slightly biased estimates (2.7–3.3%). The model calibration estimate $\hat{I}_{\mathrm{MA2}}(R)$ based on the posterior samples from ZAP+SOD or ZANB+S had a relative bias between the design-based estimate and the model calibration $\hat{I}_{\mathrm{MA1}}(R)$ estimate (0.4–1.4%). The biases from model calibration $\hat{I}_{\mathrm{MA1}}(R)$ to posterior median from ZAP+SOD were nega-

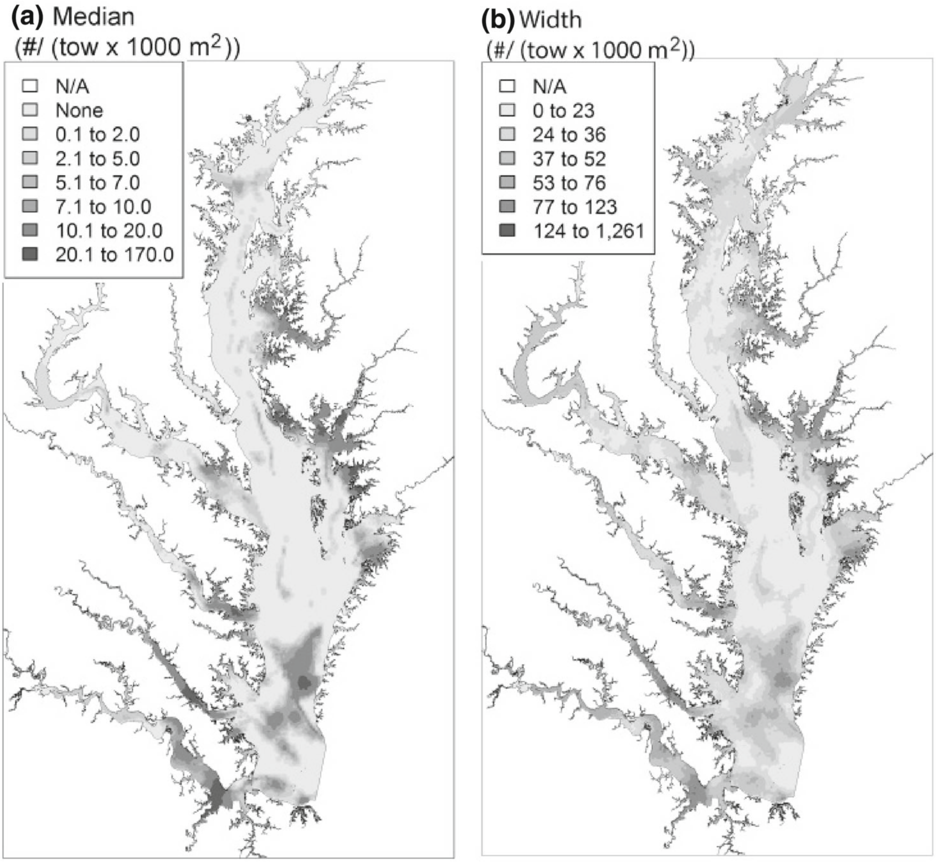

Figure 5. Maps of Chesapeake Bay during a low abundance year showing **a** blue crab winter CPUE density (number/tow/1000 m^2) and **b** inter quartile range based on spatial modeling of winter dredge survey data for winter of 2000/2001.

tive in the simulation results, while those from the real data application were mostly positive, i.e., larger than the design-based estimates.

The calibration estimators based on misspecified models (ZAP & ZANB) generated similar MSEs as the design-based estimator. Although the spatial model calibration estimates based on ZAP+SOD and ZANB+S were slightly biased, the method generated lower MSEs than those derived from stratified design-based estimates. The calibration estimator $\hat{I}_{\mathrm{MA2}}(R)$ based on the posterior samples yielded smaller MSEs than the calibration estimator $\hat{I}_{\mathrm{MA1}}(R)$ based on the posterior median in all three years. The calibration estimator $\hat{I}_{\mathrm{MA1}}(R)$ generated smaller MSEs than the design-based estimator in 2006 and 2012, but not in 2001. Overall, the calibration estimator $\hat{I}_{\mathrm{MA2}}(R)$ based on posterior samples from the correctly specified model generated smaller MSE relative to the design-based approach with more than a 15% reduction in MSE in medium and high abundance scenarios (2006 and 2012).

The estimated model variance components were smaller than the simulated values (Table 3). Specifically, the underestimation was strongest when assuming a Poisson model (ZAP). The ZANB calibration generated the smallest negative bias. The two calibrations based on spatial smoothing (ZANB+S & ZAP+SOD) generated negative bias between ZAP

Table 2. Simulated bias, mean squared error (MSE) for the average abundance index based on 720 Monte Carlo samples.

Year	Model	Relative Bias			MSE		
		HT(%)	MA_1(%)	MA_2(%)	HT	MA_1	MA_2
2001	ZAP	0.2	0.2	0.2	0.136	0.136	0.136
	ZANB	0.2	0.2	0.2	0.136	0.136	0.136
	ZANB+S	0.2	−2.9	−0.4	0.136	0.145	0.132
	ZAP+SOD	−0.1	−3.3	−0.6	0.149	0.158	0.146
2006	ZAP	−0.1	−0.1	−0.1	0.304	0.305	0.304
	ZANB	−0.1	−0.1	−0.1	0.304	0.305	0.304
	ZANB+S	−0.1	−2.7	−1.2	0.304	0.261	0.253
	ZAP+SOD	−0.1	−2.8	−1.1	0.299	0.268	0.255
2012	ZAP	−0.1	−0.1	−0.1	3.691	3.594	3.628
	ZANB	−0.1	−0.1	−0.1	3.691	3.641	3.669
	ZANB+S	−0.1	−2.7	−1.4	3.691	3.169	3.069
	ZAP+SOD	−0.3	−3.0	−1.4	3.695	3.201	3.131

HT Horvitz-Thompson estimates; MA_1 model assisted estimates calibrated to posterior median using pseudo-empirical likelihood; MA_2 model assisted estimates calibrated to posterior samples using pseudo-empirical likelihood; *ZAP* zero-augmented Poisson model; *ZANB* zero-augmented negative Binomial model; *ZANB+S* zero-augmented negative Binomial model with SPDE random effects; *ZAP+SOD* zero-augmented Poisson model with over-dispersed and SPDE random effects.

Table 3. Ratio between the simulated variance components for the average abundance index based on 720 Monte Carlo samples and the estimated variance component using a classical estimator and pseudo-empirical likelihood calibration.

Year	Model	Total			Model		Design		
		HT	MA_1	MA_2	HT	MA	HT	MA_1	MA_2
2001	ZAP	0.476	0.768	0.769	0.000	0.415	1.643	1.627	1.637
	ZANB	0.476	0.988	0.987	0.000	0.723	1.643	1.640	1.636
	ZANB+S	0.476	0.753	0.872	0.000	0.565	1.643	1.255	1.712
	ZAP+SOD	0.432	0.712	0.796	0.000	0.568	1.222	1.015	1.251
2006	ZAP	0.615	0.832	0.835	0.000	0.508	1.093	1.083	1.089
	ZANB	0.615	1.102	1.103	0.000	1.121	1.093	1.087	1.088
	ZANB+S	0.615	0.760	0.957	0.000	0.725	1.093	0.806	1.227
	ZAP+SOD	0.629	0.802	0.998	0.000	0.789	1.132	0.819	1.234
2012	ZAP	0.663	0.751	0.778	0.000	0.260	1.230	1.193	1.235
	ZANB	0.663	1.065	1.092	0.000	0.933	1.230	1.181	1.230
	ZANB+S	0.663	0.711	0.933	0.000	0.639	1.230	0.819	1.326
	ZAP+SOD	0.661	0.853	1.041	0.000	0.778	1.228	0.971	1.377

Total columns denote $\hat{V}/V$ where V and $\hat{V}$ denote the estimated and simulated total variance component. Model columns denote $\hat{V}_1/V_1$ where $\hat{V}_1$ and V_1 denote estimated and simulated model variance component. Design columns denote $\hat{V}_2/V_2$ where $\hat{V}_2$ and V_2 denote estimated and simulated design variance components. Other acronyms are the defined in caption of Table 2.

and ZANB. The model variance components were assumed to be zero in the design-based framework (Table 3).

The estimated variance components due to sampling were larger than the simulated values. Calibration based on ZAP or ZANB generated similar estimates as the Horvitz–Thompson estimator. When calibration was based on spatial and overdispersed models

(ZAP+SOD or ZANB+S), the estimator $\hat{I}_{\mathrm{MA1}}(R)$ based on posterior median generated smaller estimates than the estimator $\hat{I}_{\mathrm{MA2}}(R)$ based on the full posterior sample.

The design-based estimator generated smaller total variance estimates than simulated values in all three years. Calibration based on ZAP generated negatively biased total variance estimates. The corresponding estimator based on ZANB was almost unbiased. When calibration was based on spatial and overdispersed models (ZANB+S or ZAP+SOD), the variance estimator based on posterior median $\hat{I}_{\mathrm{MA1}}(R)$ was more negatively biased than the estimator $\hat{I}_{\mathrm{MA2}}(R)$ based on the full posterior sample. The estimated total variances based on $\hat{I}_{\mathrm{MA2}}(R)$ were close to the simulated values, except in low abundance scenarios (year 2001). The properties of model calibration based on ZANB+S did not change qualitatively when we altered the parameters from the nine chosen hyper-priors (Supp. Table 1 and 2).

5. DISCUSSION AND CONCLUSION

We proposed a Bayesian model calibration within the pseudo-empirical likelihood framework to produce improved estimates of the total variance due to varying catchability and sampling variability. We extended the geostatistical approach of Jensen and Miller (2005) with a Bayesian calibration approach. Our mean estimates were similar to that of classical geostatistical analyses, but the revised standard errors were larger than those in the original analysis. Our simulation study indicated that the standard errors from the Bayesian model calibration approach were approximately unbiased for years 2006 and 2012. Thus, the Bayesian model calibration framework provides an improved way to estimate total variance, while maintaining the approximate unbiasedness of the pseudo-empirical likelihood approach (Chen et al. 2004).

Model calibration is conventionally conducted in a likelihood framework and does not explicitly consider spatial auto-correlation (Cicchitelli and Montanari 2012; Opsomer et al. 2007). We examined both Bayesian framework and spatial auto-correlation. We found that model calibration based on a point estimate generated biased estimates. Thus, calibration based on the entire posterior sample is recommended when analyzing spatial and overdispersed survey data. This bias could have resulted from the spatially explicit modeling, which smooths the neighboring data to improve small area estimation, but likely generates bias in the global estimation (Brus and DeGruijter 1993). Fishery survey data are commonly zero-inflated and overdispersed (Liu et al. 2011; Thorson and Ward 2013). The zero-inflated model is not continuous and thus violates the assumptions for the asymptotic unbiasedness of the EL estimator (Wu and Sitter 2001). This violation could lead to the different direction of biases observed between the real data analyses and simulation studies, where the smoothed data (i.e., median of the posterior predictive distribution) were used rather than the original zero-inflated and overdispersed CPUE density in the simulation. Hence, it is important to allow for overdispersion in the working model because models without overdispersion could result in substantially biased estimates.

Multi-phase and complex survey designs are often conducted in environmental surveys and inventory studies (Opsomer et al. 2007). Although, in theory, our approach can be parameterized for a general design, in practice we are faced with the challenges regarding estimation of the inclusion probabilities, correction of informative designs (Savitsky and

Toth 2016), and implementation of the pseudo-empirical likelihood in complex survey contexts. Thus, our conclusions only apply to non-informative designs. These complex designs often collect additional covariates such as age and sex of each individual. In addition, these surveys are usually conducted over multiple years. This specific spatiotemporal information can be used to improve the specificity and accuracy of the estimates. To incorporate these factors in our approach, however, we would need a spatiotemporal working model for multivariate and zero-inflated data. Although these models are readily available in the INLA framework, they require considerable modeling efforts to code and interpret, and are much less practical for routine analyses than design-based estimates, which are independent of the underlying models and can be routinely conducted for many response variables (Opsomer et al. 2007). Hence, a more user-friendly framework, such as an R package (see online supplementary code), is needed to alleviate this limitation and make our approach more generalizable to age- and sex-specific variables collected during a sequential survey.

Our results highlight the need for improvements in the design, implementation, and analysis of surveys to reduce the variance component due to non-sampling factors. Although this component of variance is widely recognized (Dick 2004; Wilberg et al. 2010), it is usually not included in estimating the standard errors associated with indices of abundance derived from surveys. Routine collection of environmental conditions that are relevant to the population and catch process would provide important auxiliary information that can improve index estimation and reduce variance estimates. However, these auxiliary variables must be available over the entire study area. Geographic Information System and Remote Sensing technologies could be applied to generate surrogate estimates of these conditions (e.g., Bauer and Miller 2010). This information could also be used in the design phase to contextualize the spatial sampling frame (Kumar 2009), and provide improved stratification variables for the frame.

In summary, the Bayesian calibration framework based on the entire posterior sample (MA_2) provided an approximately unbiased estimator of the total variance. Application to the WDS data suggested that the approach also maintains the design-based unbiasedness of the population index estimation. The estimated CVs from Bayesian calibration were larger than those from design-based estimates and showed an increasing trend between 1994 and 2014. Simulation results suggested this increase could reflect time-varying catchability. Given that fishery-independent surveys based on probability designs are commonly used to develop time series of abundance estimates for fishery management, our proposed approach has the advantage of incorporating design and non-design factors.

ACKNOWLEDGEMENTS

We thank the Maryland Department of Natural Resources (MDNR) and the Virginia Institute of Marine Science for conducting the Chesapeake Bay blue crab winter dredge survey and G. Davis (MDNR), L. Fegley (MDNR), S. Iverson (Virginia Marine Resource Commission), and A. Sharov (MDNR) for providing the data. Two anonymous reviewers provided helpful comments. The work was funded by Grant CBSAC4 from the Chesapeake Bay Trust. This is contribution number 5383 of the University of Maryland Center for Environmental Science.

[Received October 2016. Accepted July 2017. Published Online July 2017.]

REFERENCES

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Boca Raton: CRC Press.

Bauer, L. J. and T. J. Miller. 2010. “Spatial and temporal variability in winter mortality of the blue crab (*Callinectes sapidus*) in the Chesapeake Bay”. *Estuaries and Coasts* 33:678-687

Brus, D. J., & DeGruijter, J. J. (1993). “Design-based versus model-based estimates of spatial means: Theory and application in environmental soil science”. *Environmetrics*, 4(2), 123-152.

Chen, J., & Sitter, R. R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica*, 385-406.

Chen, J., Sitter, R. R., & Wu, C. (2002). “Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys”. *Biometrika*, 89(1), 230-237.

Chen, J., Thompson, M. E., & Wu, C. (2004). “Estimation of fish abundance indices based on scientific research trawl surveys”. *Biometrics*, 60(1), 116-123.

Chesapeake Bay Program (2016). “The Data Hub. Chesapeake Bay Program”. Annapolis, Maryland, U.S. URL http://www.chesapeakebay.net/data.

Cicchitelli, G., & Montanari, G. E. (2012). “Model-assisted estimation of a spatial population mean”. *International Statistical Review*, 80(1), 111-126.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. New York: Wiley.

Dick, E.J. 2004. “Beyond “lognormal versus gamma” discrimination among error distributions for generalized linear models”. *Fisheries Research 70:351-366*.

Fieberg, J., Alexander, M., Tse, S., & St Clair, K. (2013). “Abundance estimation with sightability data: a Bayesian data augmentation approach”. *Methods in Ecology and Evolution*, 4(9), 854-864.

Horvitz, D. G., & Thompson, D. J. (1952). “A generalization of sampling without replacement from a finite universe”. *Journal of the American Statistical Association*, 47(260), 663-685.

Jensen, O. P., & Miller, T. J. (2005). “Geostatistical analysis of the abundance and winter distribution patterns of the blue crab *Callinectes sapidus* in Chesapeake Bay”. *Transactions of the American Fisheries Society*, 134(6), 1582-1598.

Kimura, D. K & Somerton, D. A. (2006).“Review of statistical aspects of survey sampling for marine fisheries”. *Reviews in Fisheries Science*. 14, 245-283.

Kumar, N. (2009). “An optimal spatial sampling design for intra-urban population exposure assessment”. *Atmospheric Environment*, 43(5), 1153-1155.

Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19).

Lindgren, F., Rue, H., & 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(4), 423-498.

Liu, H., Ciannelli, L., Decker, M. B., Ladd, C., & Chan, K.-S. (2011). “Nonparametric threshold model of zero-inflated spatio-temporal data with application to shifts in jellyfish distribution”. *Journal of Agricultural, Biological, and Environmental Statistics*, 16(2), 185-201.

Maunder, M. N., & Punt, A. E. (2004). “Standardizing catch and effort data: a review of recent approaches”. *Fisheries Research*, 70(2), 141-159.

Opsomer, J. D., Breidt, F. J., Moisen, G. G., & Kauermann, G. (2007). “Model-assisted estimation of forest resources with generalized additive models”. *Journal of the American Statistical Association*, 102(478), 400-409.

Pfeffermann, D. (2007). “Comment: struggles with survey weighting and regression modeling”. *Statistical Science*, 22(2), 179-183.

Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., & Rasbash, J. (1998). “Weighting for unequal selection probabilities in multilevel models”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1), 23-40.

Rue, H., Martino, S., & Chopin, N. (2009). “Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2), 319-392.

Särndal, C. E., Swensson, B., & Wretman, J. H. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Särndal, C. E., Thomsen, I., Hoem, J. M & Lindley, D. V. (1978).“Design-based and model-based inference in survey sampling [with discussion and reply]”. *Scandanavian Journal of Statistics*, 5(1), 27-52.

Savitsky, T. D., & Toth, D. (2016). “Bayesian estimation under informative sampling”. *Electronic Journal of Statistics*, 10(1), 1677-1708.

Sharov, A., Vølstad, J., Davis, G., Davis, B., Lipcius, R., & Montane, M. (2003). “Abundance and exploitation rate of the blue crab (*Callinectes sapidus*) in Chesapeake Bay”. *Bulletin of Marine Science*, 72(2), 543-565.

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(5), 894-903.

Thompson, S. K. (2002). *Sampling*. New York: Wiley.

Thorson, J. T., & Ward, E. J. (2013). “Accounting for space–time interactions in index standardization models”. *Fisheries Research*, 147, 426-433.

Valliant, R., Dorfman, A., & Royall, R. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. New York: Wiley-Interscience.

Wagner, T., Bence, J. R., Bremigan, M. T., Hayes, D. B., & Wilberg, M. J. (2007). “Regional trends in fish mean length at age: Components of variance and the statistical power to detect trends”. Canadian Journal of Fisheries and Aquatic Sciences, 64(7), 968–978.

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

Wu, C. (2005). “Algorithms and R codes for the pseudo empirical likelihood method in survey sampling”. *Survey Methodology*, 31(2), 239.

Wu, C., & Sitter, R. R. (2001). “A model-calibration approach to using complete auxiliary information from survey data”. *Journal of the American Statistical Association*, 96(453), 185-193.