Multi-state dead recovery mark-recovery model performance for estimating movement and mortality rates

Emily M. Liljestrand[a,*], Michael J. Wilberg[a], Amy M. Schueller[b]

[a] Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, P.O. Box 38, Solomons, MD, 20688, USA
[b] National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southeast Fisheries Science Center, 101 Pivers Island Road, Beaufort, NC, 28516, USA

ABSTRACT

Multi-state mark-recovery models are used to estimate movement and mortality rates of terrestrial and aquatic animals. These models have become especially popular in the last 20 years since technology and statistical techniques have improved to accommodate the extensive data requirements. However, the ability of multi-state mark-recovery models to estimate movement rates has received little evaluation, with few studies exploring the effects of alternative release and recovery designs on the bias and precision of estimates. Our objectives were to evaluate the effects of the spatio-temporal pattern of releases, pattern of recovery efforts, and number of releases on the performance of a multi-state mark-recovery model. We generated mark-recovery data from a spatial model and fitted them using a multi-state dead recovery model that included prior distributions on movement rates, natural mortality, and catchability. We generated data using a spatially variable schedule of releases and effort, and a release size of about 1 million individuals to mimic a mark-recovery study conducted for Atlantic Menhaden *Brevoortia tyrannus* in the late 1960s. We also ran alternative scenarios of sample size and spatially uniform releases and effort, either by themselves or in combination to determine their effects on the accuracy of the estimates. The model generally produced unbiased estimates of mortality rates with median error < 0.02 yr^{-1} for all scenarios, but some biases were present for the movement rates. Movement rates and catchability were more accurately and precisely estimated in scenarios that included spatially uniform fishing effort, while spatially uniform releases had little to no effect on bias or precision of estimated movement rates. Increased sample size improved accuracy of all parameter estimates except for the lowest movement rates. Future mark-recovery experiments that use a multi-state dead recovery model may benefit from distributing recovery effort uniformly over time and space.

1. Introduction

Multi-state mark-recovery models explicitly account for marking and recovering across multiple sites, i.e., states, which allows estimation of movement and mortality rates (Arnason, 1973; Schwarz et al., 1993). In addition to movement and mortality, multi-state mark-recovery models can estimate reproductive state, or age- and weight-specific parameters (Nichols et al., 1994; Nichols and Kendall, 1995; Sippel et al., 2015). Multi-state mark-recovery models have become more popular with the development of software such as program MARK and increased computational power (Lebreton and Cefe, 2002). Multi-state dead recovery models, which have only one tagging and one possible recovery event for each individual, are one type of mark-recovery model still in popular use (Hestbeck et al., 1991; Brownie et al., 1993; Schwarz et al., 1993). Unlike multiple resighting models, which

use detection as a proxy for survival, dead recovery models use detection to directly estimate mortality (Lindberg, 2012). Because dead recovery studies have at most two records of an individual's location (i.e. release and recovery), these models require large sample sizes to be able to detect low frequency movement events (Lindberg and Rexstad, 2002).

Identifiability among mortality, movement, and detection probability is particularly difficult in multi-state models when movement and mortality are time dependent (Kendall and Nichols, 2002; Schaub et al., 2004). Increasingly, this shortcoming has been addressed by incorporating prior probability distributions on parameters (Newman, 2000; Gimenez et al., 2007; Calvert et al., 2009). Including prior distributions has been useful when the data are limited or of poor quality, which is frequently the case when recoveries are often obtained opportunistically (Martell and Walters, 2002; Michielsens et al., 2011).

* Corresponding author.
  E-mail address: liljest3@msu.edu (E.M. Liljestrand).

Simulation studies have been used to evaluate the accuracy and performance of multi-state models (Faubet et al., 2007; Vermard et al., 2010). Data are generated using known values (e.g., mortality, catchability, and biomass), which are used to fit models that estimate these same values. The estimates are compared against the known values to determine model accuracy (e.g., Chao, 1987; Lee et al., 2011; Wetzel and Punt, 2011). For multi-state mark-recovery models, simulation studies have helped determine the optimal schedule of releases and recovery effort that reduces bias of movement rate parameter estimates (Ovaskainen et al., 2008). Spreading fishing effort across a broader spatial range can lead to more accurately estimated movement rates (Albanese et al., 2003; Yamamura et al., 2003). However, it is not yet understood how the combination of distribution of releases, recovery effort, and sample size affect the performance of a mark-recovery model that simultaneously estimates mortality and movement.

Our goal was to evaluate the accuracy and precision of a multi-state dead recovery mark-recovery model that simultaneously estimates mortality and movement. Our specific objectives were to determine whether model performance improved with 1) increased sample size, 2) spatiotemporally homogenous sampling effort, and 3) spatiotemporally homogenous releases. The mark-recovery model also used supplementary data on tag shedding/tagging mortality, detection efficiency, and fishing effort to estimate monthly movement rates among four regions, month- and region- specific catchability, and monthly natural mortality. We generated data under similar conditions as the 1966–1969 Atlantic Menhaden *Brevoortia tyrannus* mark recovery study and several alternative release and recovery scenarios (Liljestrand et al., 2018).

2. Methods

Mark-recovery data were generated with known values of movement, natural mortality, and catchability and under different scenarios of release size, spatio-temporal patterns of fishing effort, and spatio-temporal patterns of releases. The mark-recovery model was then fitted to the data to estimate movement, natural mortality, and catchability. Accuracy was assessed by comparing the estimated values to those used to generate the data. The process was repeated 500 times for each scenario. The data-generating and mark-recovery models were run in AD Model Builder (Fournier et al., 2012), and analysis of the results were conducted using R (R Core Team, 2015).

2.1. Data generating model

The data generating model generated mark-recovery data over a 42-month period and across four regions to replicate the structure of the July 1966-December 1969 data from Liljestrand et al., (2018). The initial size of the cohort upon release time, $t = T$, and region, $r = R$, was calculated as the product of the number released, $I_{T,R}$, which varied among scenarios, the region-specific tagging mortality rate, G_R, 0.1 in region 1, 0.2 in region 2, 0.25 in region 3, and 0.4 in region 4, from (Dryfoos et al., 1973),

$$N_{T,R,T,R} = I_{T,R}(1 - G_R) \tag{1}$$

We calculated the size of a tagged cohort, $N_{T,R,t,r}$, present at time, t, and region, r, that were released at time, T, in region, R, by sequentially applying mortality and movement rates to the number of individuals that survived until that time. The number of individuals before applying movement, $N^*_{T,R,t,r}$, was calculated as the product of abundance and survival from monthly fishing and natural mortality,

$$N^*_{T,R,t,r} = N_{T,R,t,r}e^{-M_t - q_{t,r}E_{t,r}} \tag{2}$$

Natural mortality, M_t, was calculated as the product of a median value of 0.10 mo^{-1}, approximately the value estimated in Liljestrand et al., (2018), and a lognormal process error, δ_t with a log-scale mean of zero and standard deviation of 0.2,

$$M_t = 0.1 * e^{\delta_t} \tag{3}$$

Fishing mortality was the product of month- and region-specific catchability, $q_{t,r}$, and effort, $E_{t,r}$, which depended on the scenario. Catchability included a multiplicative process error, γ_t (same process error across regions), applied to the on the month- and region- specific catchability values, $q_{m,r}$, that was lognormally distributed with a log-scale mean of zero and standard deviation of 0.2,

$$q_{t,r} = q_{m,r} * e^{\gamma_t} \tag{4}$$

The process errors in natural mortality and catchability were included so that the estimation model was different from the simulation model that generated the data sets. The number of individuals after movement occurred, $N_{T,R,t+1}$, was calculated as the product of the monthly movement rate matrix, φ_t, (Table A.1), and the vector of individuals in each region, $N^*_{T,R,t}$,

$$N_{T,R,t+1} = \varphi_t N^*_{T,R,t} \tag{5}$$

No movement occurred between November to May, i.e., φ_t was an identity matrix.

The number of observed recoveries from each release event was generated using a binomial distribution with the number of individuals caught in a region, $A_{T,R,t,r}$, and the time- and region-specific detection efficiency, $\varepsilon_{t,r}$,

$$C_{T,R,t,r} \sim Binom(A_{T,R,t,r}, \varepsilon_{t,r}) \tag{6}$$

The expected number of individuals caught in a region was calculated as the product of the probability of being captured (based on the Baranov catch equation) and the number of individuals in that region,

$$A_{T,R,t,r} = \frac{q_{t,r}E_{t,r}}{M_t + q_{t,r}E_{t,r}}(1 - e^{-M_t - q_{t,r}E_{t,r}})N_{T,R,t,r} \tag{7}$$

The model lacked age- or size- structure and, therefore, assumed catchability, natural mortality, and movement probability was the same for all individuals in a given month and region.

2.2. Estimation model

Parameters were estimated by fitting a mark-recovery model to the generated data sets. The mark-recovery estimation model was a multi-state version of the Brownie dead recovery model that included prior probabilities on movement rates, instantaneous natural mortality, and catchability parameters (Hoenig et al., 1998; Liljestrand et al., 2018). The estimation model had 42 monthly time steps and tracked the abundance over time and across four regions based on the movement rate and natural and fishing mortality using the same equations as the data generating model. The likelihood for the observed data was the negative binomial distribution. A weakly informative Dirichlet distribution was used as a prior on the movement rates, and the values of the prior were calculated using a one-dimensional diffusion model that assumed individuals traveled randomly in a north-south direction over the course of a month. The proportion of individuals that moved to an adjacent region in one month after being released from the center of a region was modeled as a normal distribution with mean 0 and standard deviation, σ_w. The standard deviation was calculated from an average swimming speed, 2.09 km/hr, and number of days in a month (i.e., the distance traveled under directed swimming). The mean prior probability of moving between regions was calculated by integrating the probability density over the distance between the center of starting region and the ending region's northern and southern boundaries. An uninformative uniform distribution was placed on the log-scale natural mortality rate and catchability was conditioned such that deviations from the mean on the log scale were normally distributed. Tag shedding and mortality, detection efficiency, and the overdispersion value of the negative binomial distribution were assumed to be known without error. For the simulation study, we did not use Markov Chain Monte

Carlo to estimate the parameters because of time constraints. Rather, we approximated the full Bayesian model by estimating the highest posterior density estimates. Results from data sets with failed model convergence were removed from the analysis. Models failed to converge if the maximum gradient component was > 0.0001 and the Hessian did not invert.

2.3. Model performance

Error was calculated as the difference between the estimated parameter value and the true value, and relative error was calculated as the error divided by the true value,

$$relative = \frac{estimate - true}{true} \tag{8}$$

We characterized bias as the median difference between the estimated and true value and precision as the standard deviation of the estimates.

The root mean squared error, *RMSE*, was calculated as a measure of accuracy for each estimate,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{500}(estimate_i - true)^2}{500}} \tag{9}$$

The *RMSE* was also calculated over all movement estimates (each model estimated 96 movement rates),

$$RMSE_\varphi = \sqrt{\frac{\sum_{i=1}^{500}\sum_{p=1}^{96}(estimate_{i,p} - true_p)^2}{500*96}} \tag{10}$$

The *RMSE* was calculated for all catchability estimates, where *P* indicated the total number of catchability parameters: $P = 48$ for uniform effort scenarios and $P = 33$ for variable effort scenarios,

$$RMSE_q = \sqrt{\frac{\sum_{i=1}^{500}\sum_{p=1}^{P}(estimate_{i,p} - true_p)^2}{500*P}} \tag{11}$$

2.4. Data generating scenarios

We generated 500 mark-recovery data sets for each of the 12 scenarios. The scenarios differed in their release size, spatial and temporal distribution of effort, or the spatial and temporal distribution of releases (Table 1). In the base model (Base), the release size, effort structure, and release structure were set to the same conditions as in the Atlantic

Table 1
Data generating scenario with descriptions of release size, and release and effort structure. Variable effort structure was spatiotemporally heterogeneous according to the 1966–1969 Atlantic Menhaden fishery, and uniform effort was set to 50 vessel weeks in every time and region. Variable release structure was spatiotemporally heterogeneous according to the 1966–1969 Atlantic Menhaden mark-recovery study, and uniform release structure had 5,000 releases in every time and region.

Model	Release Size	Effort Structure	Release Structure
500,000	533,179	Variable	Variable
500,000 + Effort	533,179	Uniform	Variable
500,000 + Release	533,148	Variable	Uniform
500,000 + Effort + Release	533,148	Uniform	Uniform
Base	1,066,357	Variable	Variable
Effort	1,066,357	Uniform	Variable
Release	1,066,296	Variable	Uniform
Effort + Release	1,066,296	Uniform	Uniform
2,000,000	2,132,714	Variable	Variable
2,000,000 + Effort	2,132,714	Uniform	Variable
2,000,000 + Release	2,132,592	Variable	Uniform
2,000,000 + Effort + Release	2,132,592	Uniform	Uniform

Menhaden mark-recovery model (Table 1; Table A.2-Table A.5). A total of 1,066,357 individuals were released across 42 months, to match the Atlantic Menhaden mark-recovery study (Liljestrand et al., 2018). The base fishing effort matched the values of Atlantic Menhaden reduction fishery during July 1966-December 1969, and, for confidentiality reasons, cannot be reproduced in its entirety. A summary by year is available in Table A.6. Effort was greatest in region 2, about half as much as region 2 in regions 1 and 3, and about 10% as much as region 2 in region 4. Effort was low or zero for all regions between November-April. Median catchability, $q_{t,r}$, was set to the values estimated in Liljestrand et al., 2018, ranging from 1.47×10^{-4} to 0.013 and averaging 0.0025 vessel^{-1} week^{-1}. The average catchability in region 2 was about 10% of the other three regions. Catchability was highest between July-September. Time- and region- specific detection efficiency, $\varepsilon_{t,r}$, was set to the values estimated in Liljestrand et al., 2018. The detection efficiency ranged from 0.65-0.78 in region 1, from 0.34-0.57 in region 2, from 0.68-0.85 in region 3, and 0.62-0.67 in region 4.

The effect of release size was evaluated by scaling the number of releases by 0.5 (500 K) or 2.0 (2mil), relative to the base scenario to 533,179 or 2,132,714 individuals, respectively. The effect of spatio-temporally uniform fishing effort (Effort) was tested by setting the effort, $E_{t,r}$ to 28 vessel weeks for every time step and region. Two additional changes were made to the (Effort) models to reflect fishing effort occurring during all months and regions. The median catchability, $q_{t,r}$, was set to 0.0025 for every month and region. Detection efficiency, $\varepsilon_{t,r}$, was set to 0.67 in region 1, 0.47 in region 2, 0.82 in region 3, and 0.61 in region 4 for every time step. The effect of spatio-temporally uniform releases (Releases) was tested by setting the number of releases, $I_{t,r}$ to 6347 individuals in every time step and region to match the overall number of releases in the base scenario.

To assess the ability to accurately estimate natural mortality during the 1966–1969 Atlantic Menhaden tagging study, additional data-generation scenarios were also conducted by setting the monthly natural mortality to 0.15 mo^{-1} or 0.05 mo^{-1} under the base release size, effort structure, and release structure. To assess the appropriateness of approximating the full Bayesian model with highest posterior density estimates, the first 25 mark-recovery data sets generated under the base scenario were analyzed using full Bayesian integration with Markov Chain Monte Carlo using Metropolis- Hastings algorithm (Fournier et al., 2012). The medians of the posterior distributions were compared to the highest posterior density estimates.

3. Results

The estimation models failed to converge (maximum gradient component > 0.0001) for some data sets, and the parameter estimates from these models were removed from subsequent analyses. In all scenarios, the model failed to converge for fewer than 2% of the data sets. The parameter estimates from the full Bayesian estimation were very similar to the highest posterior density estimates. The difference between the median of the posterior distribution and the highest posterior density estimate did not exceed ± 0.02 (14%) for monthly natural mortality, ± 0.0002 (7%) for catchability, and ± 0.006 for movement rates.

Performance of the mark-recovery model differed under alternative data generating scenarios (Fig. 1). It produced accurate estimates of most of the parameters, although some parameters had consistent bias. Scenarios that had uniformly distributed effort performed best compared to the base scenario. Increasing the release size led to improvements in accuracy, decreasing bias and improving precision, but spatiotemporally uniform releases had little to no effect on performance.

Scenarios with uniform effort had lower RMSEs for movement and catchability parameters than scenarios with a uniform distribution of releases or the base scenario (Fig. 1). Adding uniform effort to a scenario decreased RMSE of movement parameters by 9–33%. For all scenarios, adding uniform effort decreased RMSE of catchability

a)

b)
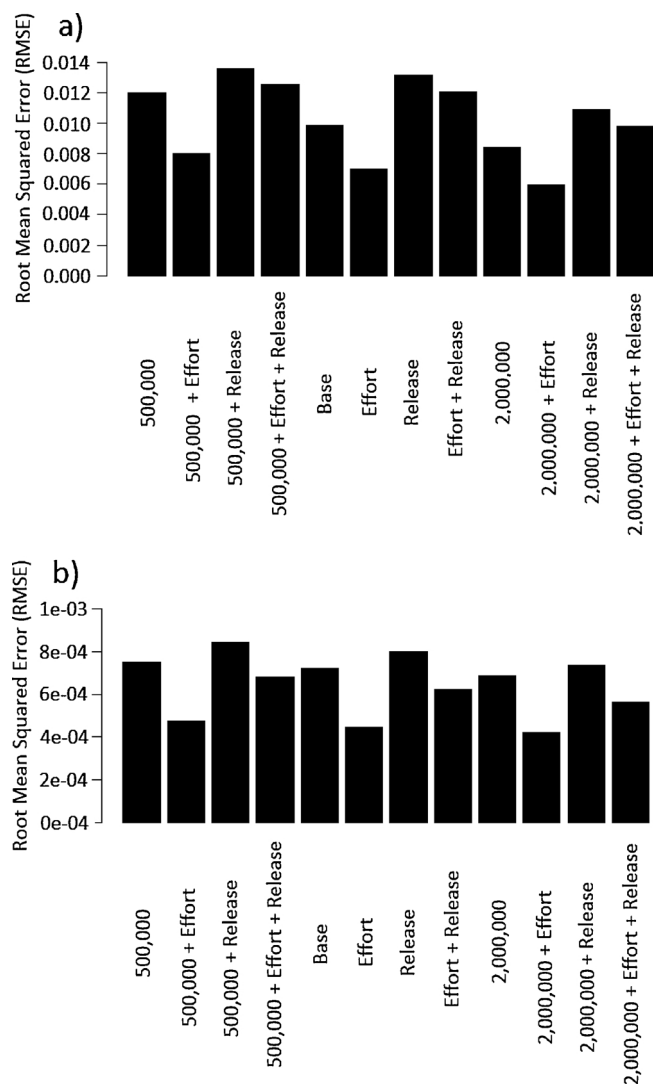
Fig. 1. Total root mean squared error (RMSE) of estimated parameters of a) movement and b) catchability by scenario. A lower RMSE value indicates more accurate estimates.

parameters by 19–38%. Both uniform effort and releases together increased the RMSEs of movement parameters by 5–22% and decreased the RMSEs of catchability parameters by 9–18% compared to the base. For all parameters and scenarios, doubling the release size decreased the RMSEs by 2–18%.

Movement rates were accurately estimated across scenarios, but were more accurate in scenarios with spatiotemporally uniform effort and increased releases (Figs. 2–5; Figs. A.1-A.20). The largest median relative error values (> 100% or < -100%) occurred for the smallest movement rates (< 0.005). Movement rates with true values > 0.05 were estimated fairly accurately with median relative errors < 23%, and movement rates with true values > 0.1 had median relative errors < 12%. In all the scenarios with the same number of releases as the base scenario, several parameters with true values < 0.005, such as those between non-adjacent regions, had median relative errors larger than 100%. Including uniform effort increased precision for some larger movement rates as well as decreasing bias (Figs. A.1-A.12). Increasing the number of released individuals increased the precision and decreased bias of monthly movement rate estimates with one exception (Figs. A.7-A.11); the estimated movement rates between regions 1 and 4 became more negatively biased with increased releases. Doubling the release size increased the median relative error of estimated movement between regions 4 and 1 by 0.6–23%. However, the median relative

error was less than 100%, despite the small values of these movement rates (10^{-7} - 10^{-6}) (Figs. A.7-A.11). The median error was within ± 0.08 for all monthly movement rates between May and October for all variable effort and release scenarios (Figs. 2 and 3; Figs. A.13-A.20). The scenario that included both uniform effort and releases decreased the median error of estimated movement rates by a small amount relative to the base scenario (average of 0.001). The scenario with uniform effort alone reduced the bias of the movement rate estimates. The scenarios with uniform releases did not notably improve the accuracy of movement rates compared to the base scenario. Estimated movement rates under scenarios with uniform releases were less precise than under scenarios with variable releases.

In the base scenario, the half-year movement rates had a 16% greater average median relative error compared to the monthly movement rates. Estimated movement to region 1 was positively biased, and movement to region 3 was negatively biased (Figs. 4 and 5). Scenarios that included uniform effort and releases and scenarios with increased release size increased the accuracy of most half-year movement rates relative to the base (Figs. 4 and 5). The average median relative error for half-year movement rates in the base scenario was 6%, in the uniform effort scenario was -20%, in the uniform release scenario was -6%, in the uniform effort and release scenario was -25%, and in the 2Mil scenario was -0.5%. The larger half-year movement rates (> 0.01) had higher error, but lower relative error compared to the smaller (< 0.01) half-year movement rates (Figs. 4 and 5; Fig. A.6; Fig. A.12). The half-year movement rates were less precise and more biased compared to monthly movement rates.

Natural mortality was estimated accurately and precisely with little bias in all scenarios (Fig. 6). The median relative error of the estimated natural mortality rate across all scenarios was approximately 7–19%, indicating that the model was positively biased for this parameter. The estimates of natural mortality became less biased with increased sample size. Scenarios that included uniform effort and releases, by themselves or in combination, did not reduce bias in estimates of natural mortality.

The patterns of error in the estimated fishing mortality rate were similar across scenarios (Fig. 7). Estimates were slightly positively biased in regions 1, 2, and 3. The smallest median error was in region 4 for the base and uniform release scenario. The smallest median error was in region 3 for the uniform effort scenario, and in region 1 for the uniform release and effort scenario. Fishing mortality rates (per year) were estimated to within 0.70 yr^{-1} of the true value in variable effort scenarios and 0.40 yr^{-1} in uniform effort scenarios. The true value of fishing mortality varied from 0.09 yr^{-1} to 2.16 yr^{-1} across regions under variable effort scenarios and varied from 0.68 yr^{-1} to 1.05 yr^{-1} in scenarios with uniform effort, depending on the process error. In no scenario or region did the median error of the average annual fishing mortality exceed 0.25 yr^{-1}. The scenario with uniform effort most strongly changed the error relative to the base scenario in region 3 where the median error decreased by 0.01–0.14 yr^{-1}. Adding uniform effort to the base scenario increased the median error in regions 1, 2, and 4, up to 0.18 yr^{-1}. Adding uniform effort to the base scenario decreased the precision in regions 1, 2, and 4 while increasing the precision in region 3. Doubling the release size marginally decreased the median bias and improved precision of yearly fishing mortality in all regions and years (Fig. A.21).

4. Discussion

Our multi-state mark-recovery model estimated movement and mortality rates precisely and accurately. As expected, accuracy increased when fishing effort was uniformly distributed over space and time (Ovaskainen et al., 2008). Additionally, a larger number of releases generally resulted in more accurate estimates. For the monthly movement rates, the largest values, > 0.1, had the largest range of error values, though the median relative error was less than 12%. Uniform releases alone did not improve estimates of movement probabilities or
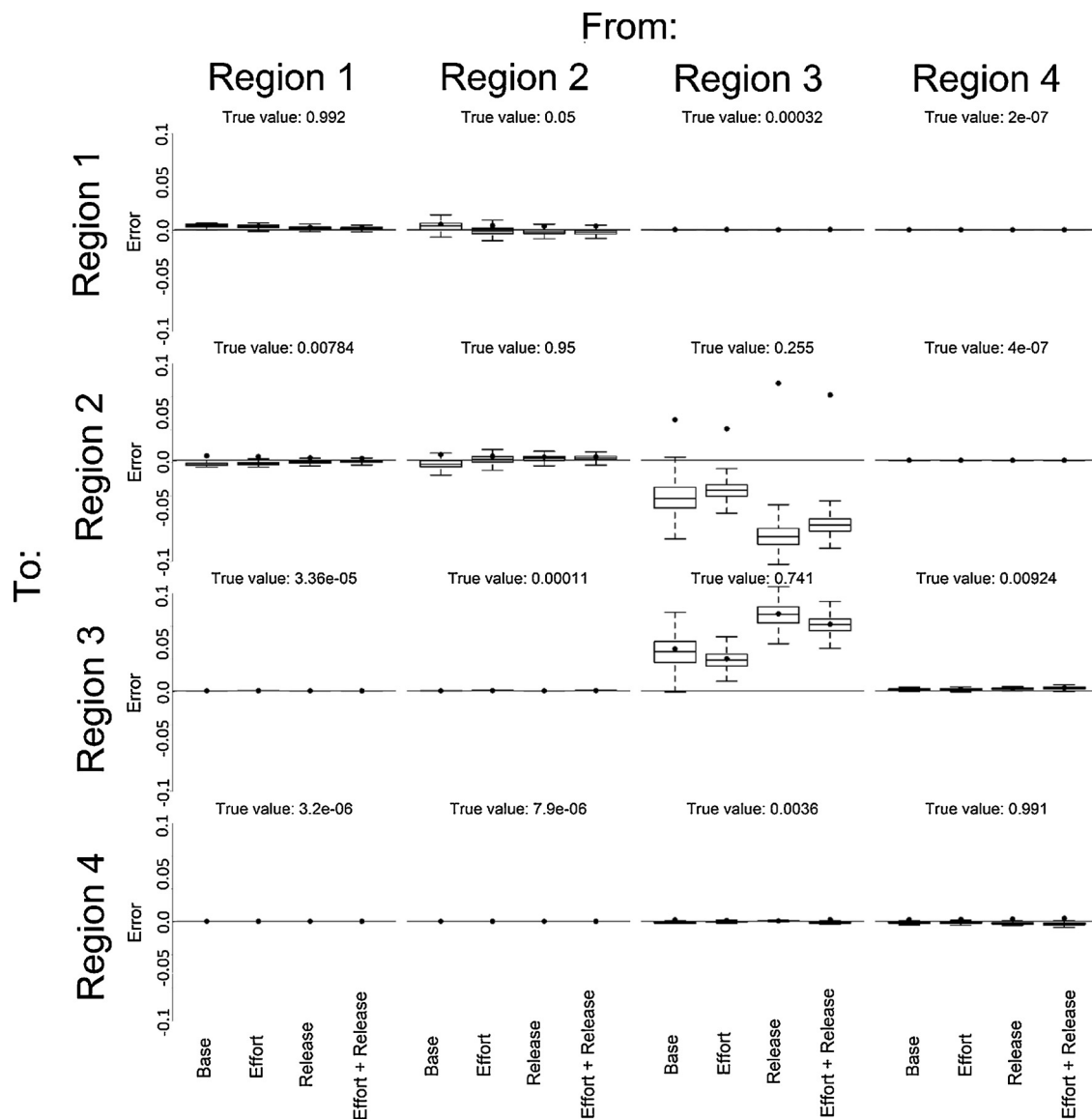
Fig. 2. Box plots of errors (estimate-true value) of monthly movement rates from June to July for four scenarios of effort and release structure. See Table 1 for scenario definitions. The boxes represent the 25% and 75% quantiles, and the line in the center indicates the median value. Whiskers indicate the 2.5 and 97.5% quantiles of error values, and the points indicate the root mean squared error.

fishing mortality rates, likely because sufficient sampling is required in all regions and times to detect movements. Additionally, spatially- and temporally- uniform releases decreased the number of releases in most months from region 3, which was an area of high sampling under the base scenario. The estimated natural mortality rate was nearly unbiased in all scenarios, and the bias and RMSE decreased by 20–24% when the release size was doubled. The model estimated natural mortality well regardless of the true value. In simulated stock assessment models, natural mortality estimates were slightly positively biased for the majority of species studied (Lee et al., 2011). Mortality rates can also become positively biased if there is emigration out of the system (Pine et al., 2011). Some positive bias might also be expected because the uniform prior placed on the log scale natural mortality, U(-5,5) was larger than the true log scale value, -1. Additionally, the priors on the other parameters may have contributed to the bias in the estimated natural mortality.

Movement rates were most accurately estimated when uniform releases were combined with uniform recovery effort, although the release structure had less of an effect than effort. A previous study also found that making recovery effort more spatially- and temporally-

uniform can improve the accuracy of movement estimates (Ovaskainen et al., 2008). Conventional tags can underestimate movement distances when effort is not well distributed spatially, compared to electronic tags, likely because conventional tags rely directly on the fishery, rather than electronic monitors for detections (Bolle et al., 2005). Similarly, the lower movement rates (i.e., the farther distances) may also be poorly estimated if there is insufficient effort on the extremes of a stock's range. High intensity trapping in close proximity to the release area can negatively bias movement rate estimates by removing those individuals who might otherwise have undergone movement, essentially decreasing the sample size (Yamamura et al., 2003).

The model was unable to accurately estimate very low movement probabilities. When the true movement probability was < 0.001, median relative errors often exceeded 300% (Figs. A.1-A.6). A much larger number of releases than used in our study is likely necessary to accurately estimate these low movement probabilities. For example, the true movement rate between regions 1 and 4 did not exceed 10^{-5}, which would require at least 100,000 marked individuals in each region to expect at least one to move to the other region. Furthermore, low fishing effort in these regions would exacerbate the problem because
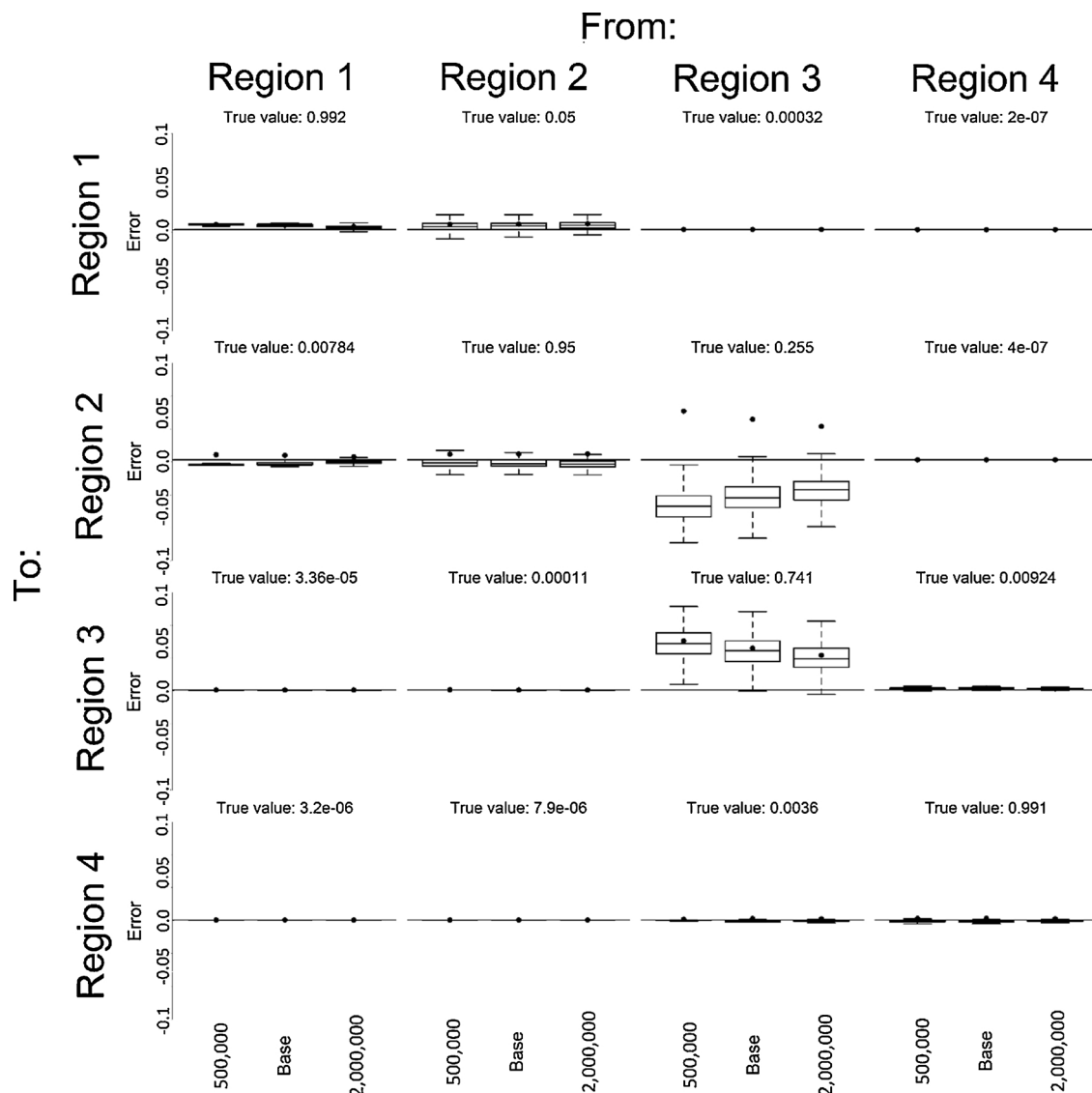
Fig. 3. Box plots of errors (estimate-true value) of monthly movement parameters from June to July for three scenarios of release size. See Table 1 for scenario definitions. Box plot definitions are the same as Fig. 2.

individuals that did move would be unlikely to be captured. Making the effort more uniform increased effort in under-fished regions, and consequently, increased the accuracy of estimates of low movement rates in the range of 10^{-5} to 10^{-3} (such as movement from region 3 to 1 or region 4 to 2). Movement in the opposite direction, from region 1 to 3 or region 2 to 4, respectively, did not see quite the same improvement with uniform effort, likely because regions 3 and 4 already had high fishing effort.

Very low movement rates among subpopulations, which are critical for understanding speciation, biodiversity, and conservation, are often detected using genetic data rather than mark-recovery information (Hey and Nielsen, 2004). One of the original objectives of the 1966–1969 Atlantic Menhaden mark-recovery study was to determine if there were two populations separated by Cape Hatteras, a series of islands in the center of North Carolina's coast (Dryfoos et al., 1973). Fish tagged north of Cape Hatteras were recovered with high frequency in the south and vice versa, which confirmed that a single coast-wide population does exist, a finding that has since been supported by genetic studies (Anderson, 2007; Lynch et al., 2010). However, under the base release size of about 1 million individuals, movement rates < 0.001 were estimated with more than 125% median relative error, which increased to more than 175% in the uniform effort scenario.

Therefore, models similar to the one we tested are incapable of reliably detecting "one individual every other generation" that keeps relatively isolated populations genetically coherent (Wright, 1931). However, the median absolute error did not exceed 0.04 yr^{-1} in the base model, which for many applications of the mark-recovery model would be acceptable accuracy.

The spatial scale for the simulating and estimating models mimicked the four regions designated in the 1966–1969 Atlantic Menhaden *Brevoortia tyrannus* mark recovery study. We assumed that all released individuals within a month and region had the same probabilities of movement and mortality within their first month. Realistically, releases were in multiple locations with some closer to regional boundaries, which may have increased the movement rate into adjacent regions. Additional studies could explore these dynamics by using a simulation model with a finer spatial and temporal scale than the estimation model.

Although making the releases and fishing effort more spatially and temporally uniform increased the accuracy and precision of the movement rate estimates in our study, it is often impractical to implement such a schedule in a real study. The Atlantic Menhaden fishery did not continue after October in the New England, New York, or New Jersey regions because the conditions became unfavorable for fishing
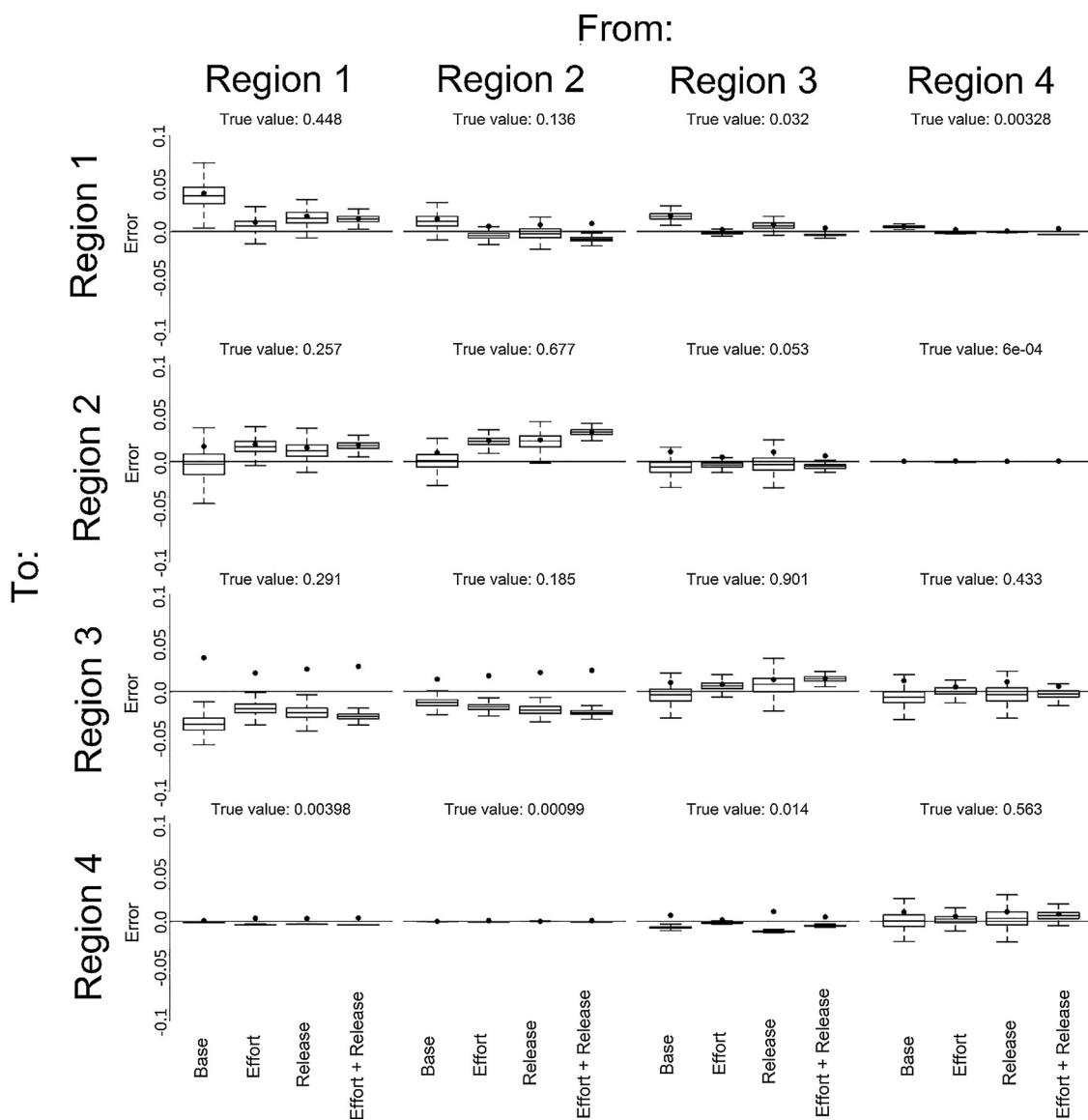
Fig. 4. Box plots of errors (estimate–true value) of half year movement parameters from October to May for four scenarios of effort and release structure. See Table 1 for scenario definitions. Box plot definitions are the same as Fig. 2.

and the fish more difficult to locate (Bigelow and Schroeder, 1953; June, 1961). Although increasing effort to 50 vessel weeks during the winter in the northernmost regions would help to accurately estimate Atlantic Menhaden movement and natural mortality, the recoveries are dependent on the reduction fishery and therefore constrained by their patterns of operation. Getting the broad spatial range of recoveries necessary to improve the movements of this mark-recovery model would likely necessitate strong participation from the fishery including fishing at times when they otherwise would not.

The number of tagged individuals should be scaled according to the goals of the study, the fishing mortality rates, and the detection probabilities. We simulated a very large number of releases to replicate the Atlantic Menhaden mark-recovery study from the 1960s. Accuracy generally increased with increased number of releases, and the scenarios with 2 million released individuals produced the most accurate and precise estimates of movement, natural mortality, and fishing mortality rates. The accuracy was still reasonably high when 1 million individuals were released, but decreased substantially for some movement rates when the releases were decreased to 500,000. Large-scale studies are possible, but require substantial planning and working with partners to implement. For example, from 1979 to 2009 over 300,000

Alaskan sablefish *Anoplopoma fimbria* were tagged by NMFS and the Alaska Department of Fish and Game in cooperation with the Japanese government (Hanselman et al., 2015). The U.S. Geological Survey has an ongoing bird banding study that has tagged nearly 40 million individuals across half a century, but most short-term mark-recovery studies do not come close to a sample size of 530,000, the lowest release size across all our scenarios. Thus, studies that rely on dead recovery approaches to estimate movement appear to require large sample sizes. Dead recovery studies with sample sizes smaller than we simulated will likely not be able to estimate movement rates accurately, unless the geographic regions are defined such that there is a substantial (> 0.1) probability of moving between those regions for each time step. Movement rates > 0.1 were accurately estimated and largely unaffected by reductions in release size.

Dead recovery models can be used to accurately estimate movement and mortality if there is sufficient sampling and supplementary information on tag shedding and tag reporting or detection, which is frequently confounded with natural mortality (Hilborn, 1990). Additionally, an effective mark-recovery study that estimates movement rates should have releases, recoveries, and quantitative fishing effort data in all regions and times and for the full range of the species or else
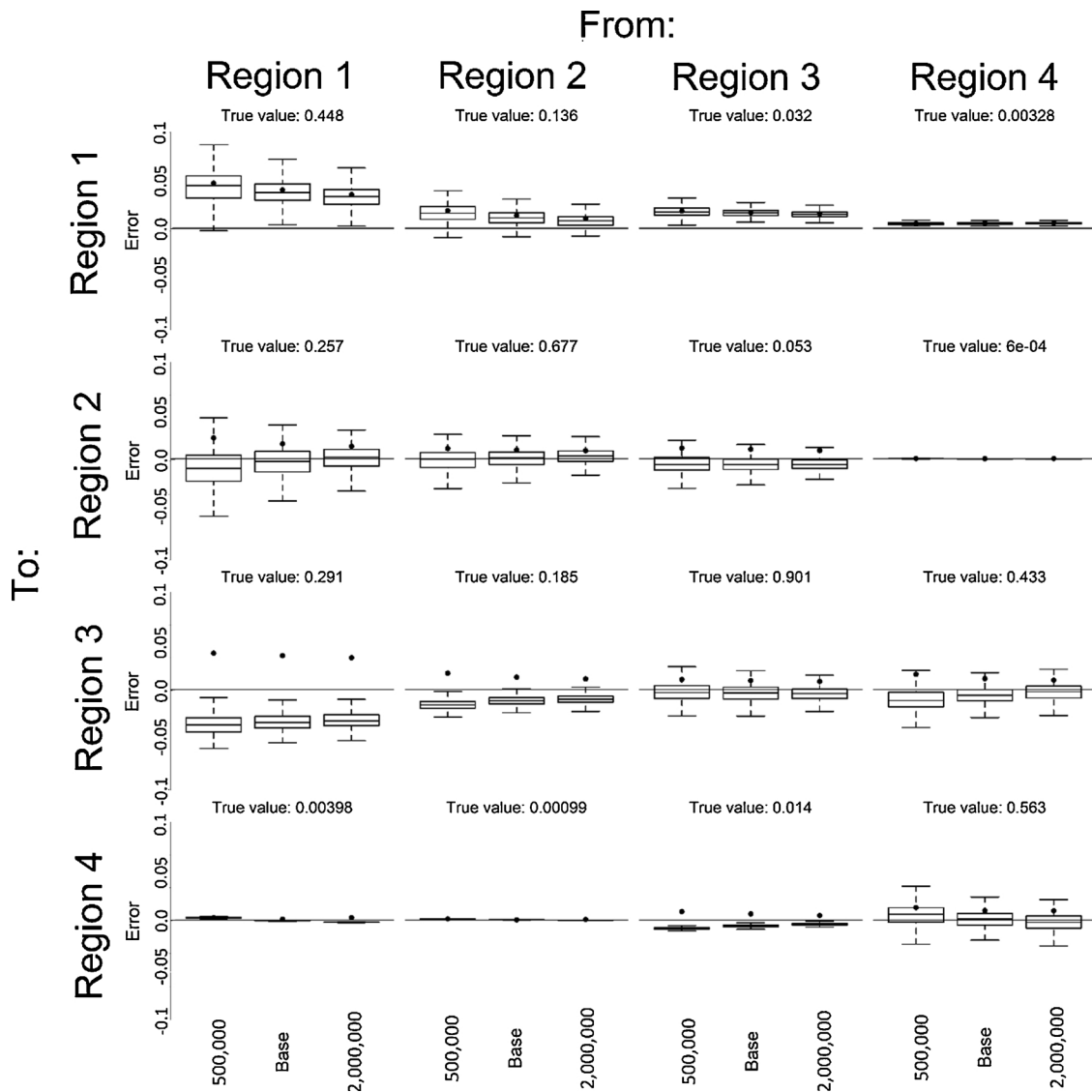
Fig. 5. Box plots of errors (estimate-true value) of half year movement parameters from October to May for three scenarios of release size. See Table 1 for scenario definitions. Box plot definitions are the same as Fig. 2.
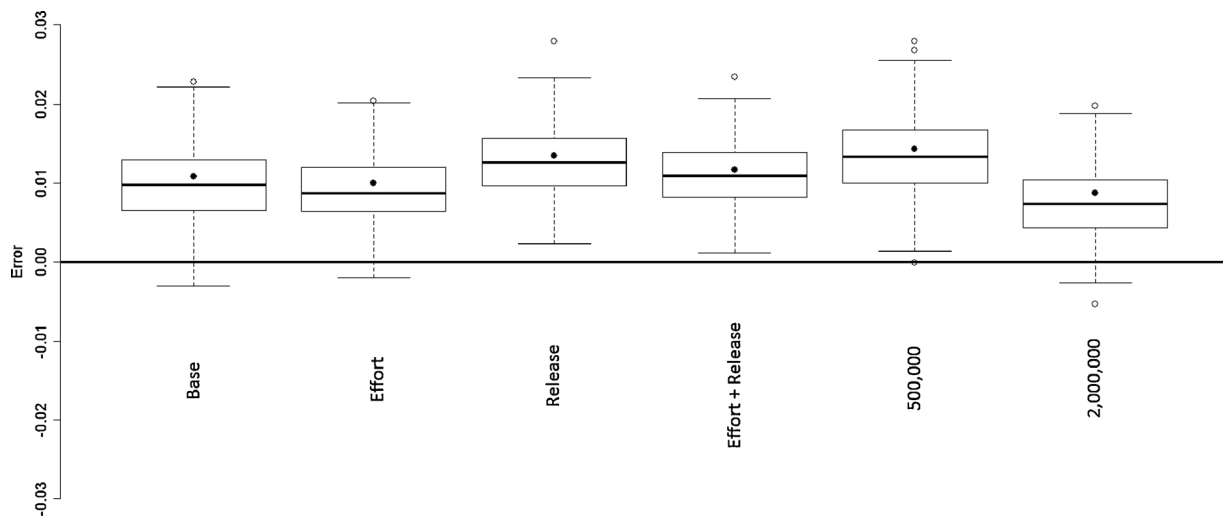

Fig. 6. Box plots of errors (estimate-true value) of monthly natural mortality parameters for six scenarios of release size, and effort and release structure. See Table 1 for scenario definitions. Box plot definitions are the same as Fig. 2.
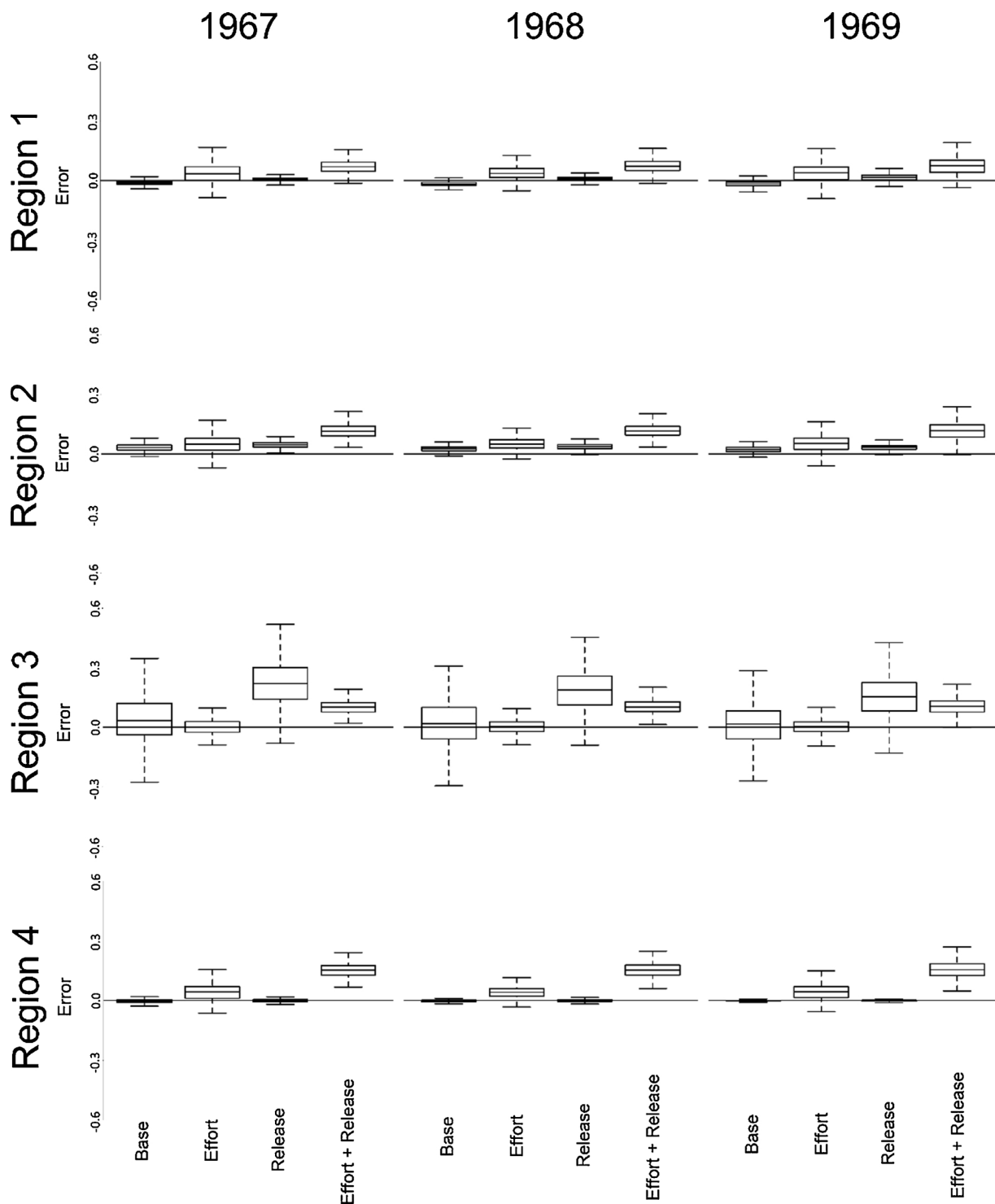
Fig. 7. Box plots of errors (estimate-true value) of yearly fishing mortality by region and year for four scenarios of effort and release structure. See Table 1 for scenario definitions. Box plot definitions are the same as Fig. 2.

movement to regions of low sampling may be misestimated (Hilborn, 1990; Schwarz et al., 1993). Bias caused by low sampling may be present in our movement and mortality estimates from Liljestrand et al. (2018), although we accounted for the lack of fishing in all regions during the late fall and winter by estimating all movement between October and May as a single event. Multi-state dead recovery models similar to the one we tested in this study have the ability to reliably estimate mortality and movement rates if sample sizes are large enough and releases and recoveries are conducted in a well-designed manner.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fishres.2018.10.014.

References

Albanese, B., Angermeier, P.L., Gowan, C., 2003. Designing mark-recapture studies to reduce effects of distance weighting on movement distance distributions of stream fishes. Trans. Am. Fish. Soc. 132, 925–939.

Anderson, J.D., 2007. Systematics of the North American menhadens: molecular evolutionary reconstructions in the genus *Brevoortia* (Clupeiformes: Clupeidae). Fish. Bull. 205, 368–378.

Arnason, A.N., 1973. The estimation of population size, migration rates and survival in a stratified population. Res. Popul. Ecol. 15, 1–8.

Bigelow, H.B., Schroeder, W.C., 1953. Fishes of the Gulf of Maine. Fish. Bull. Fish Wildl. Serv. 53, 113–118.

Bolle, L.J., Hunter, E., Rijnsdorp, A.D., Pastoors, M.A., Metcalfe, J.D., Reynolds, J.D., 2005. Do tagging experiments tell the truth? Using electronic tags to evaluate conventional tagging data. ICES J. Mar. Sci. 62, 236–246.

Brownie, C., Hines, J.E., Nichols, J.D., Pollock, K.H., Hestbeck, J.B., 1993. Analysis of multiple capture-recapture data using band-recovery methods. Biometrics 49, 1173–1187.

Calvert, A.M., Bonner, S.J., Jonsen, I.D., Flemming, J.M., Walde, S.J., Taylor, P.D., 2009. A hierarchical Bayesian approach to multi-state mark–recapture: simulations and applications. J. Appl. Ecol. 46, 610–620.

Chao, A., 1987. Estimating the population size for capture-recapture data with unequal catchability. Biometrics 43, 783–791.

Dryfoos, R.L., Cheek, R.P., Kroger, R.L., 1973. Preliminary analysis of Atlantic menhaden, *Brevoortia tyrannus*, migrations, population structure, survival and exploitation rates, and availability as indicated from tag returns. Fish. Bull. 71, 719–734.

Faubet, P., Waples, R.S., Gaggiotti, O.E., 2007. Evaluating the performance of a multi-locus Bayesian method for the estimation of migration rates. Mol. Ecol. 16, 1149–1166.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., Sibert, J., 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. 27, 233–249.

Gimenez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H., Villa, J., Pradel, R., 2007. State-space modelling of data on marked individuals. Ecol. Modell. 206, 431–438.

Hanselman, D.H., Heifetz, J., Echave, K.B., Dressel, S.C., 2015. Move it or lose it: movement and mortality of sablefish tagged in Alaska. Can. J. Fish. Aquat. Sci. 72, 238–251.

Hestbeck, J.B., Nichols, J.D., Malecki, R.A., 1991. Estimates of movement and site fidelity using mark-resight data of wintering Canada geese. Ecology 72, 523–533.

Hey, J., Nielsen, R., 2004. Multilocus methods for estimating population sizes, migration rates, and divergence time with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167, 747–760.

Hilborn, R., 1990. Determination of fish movement patterns from tag recoveries using maximum likelihood estimators. Can. J. Fish. Aquat. Sci. 47, 635–643.

Hoenig, J.M., Barrowman, N.J., Hearn, W.S., Pollock, K.H., 1998. Multiyear tagging studies incorporating fishing effort data. Can. J. Fish. Aquat. Sci. 55, 1466–1476.

June, F.C., 1961. The menhaden fishery of the United States. U.S Fish Wildl. Serv. Bur. Commer. Fish. Fish. Leaft. 521, 13.

Kendall, W.L., Nichols, J.D., 2002. Estimating state-transition probabilities for unobservable states using capture-recapture/resighting data. Ecology 83, 3276–3284.

Lebreton, J.D., Cefe, R.P., 2002. Multistate recapture models: modelling incomplete individual histories. J. Appl. Stat. 29, 353–369.

Lee, H., Maunder, M.N., Piner, K.R., Methot, R.D., 2011. Estimating natural mortality within a fisheries stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. Fish. Res. 109, 89–94.

Liljestrand, E.M., Wilberg, M.J., Schueller, A.M., 2018. Estimation of movement and mortality of Atlantic menhaden during 1966-1969 using a Bayesian multi-state mark-recapture model. Fish. Res In press.

Lindberg, M.S., Rexstad, E.A., 2002. Capture-recapture sampling designs. Encycl. Environ. 1, 251–262.

Lindberg, M.S., 2012. A review of designs for capture-mark-recapture studies in discrete time. J. Ornithol. 152, S355–S370.

Lynch, A.J., McDowell, J.R., Graves, J.E., 2010. A molecular genetic investigation of the population structure of Atlantic menhaden (*Brevoortia tyrannus*). Fish. Bull. 108, 87–97.

Martell, S.J.D., Walters, C.J., 2002. Implementing harvest rate objectives by directly monitoring exploitation rates and estimating changes in catchability. Bull. Mar. Sci. 70, 695–713.

Michielsens, C.G.J., McAllister, M.K., Kuikka, S., Pakarinen, T., Karlsson, L., Romakkaniemi, A., Pera, I., Mantyniemi, S., 2011. A Bayesian state-space mark-recapture model to estimate exploitation rates in mixed-stock fisheries. Can. J. Fish. Aquat. Sci. 63, 321–334.

Newman, K.B., 2000. Hierarchic modeling of salmon harvest and migration. J. Agric. Biol. Environ. Stat. 5, 430–455.

Nichols, J.D., Hines, J.E., Pollock, K.H., Hinz, R.L., Link, W.A., 1994. Estimating breeding proportions and testing hypotheses about costs of reproduction with capture-recapture data. Ecology 75, 2052–2065.

Nichols, J.D., Kendall, W.L., 1995. The use of multi-state capture-recapture models to address questions in evolutionary ecology. J. Appl. Stat. 22, 835–846.

Ovaskainen, O., Rekola, H., Meyke, E., Arjas, E., 2008. Bayesian methods for analyzing movements in heterogeneous landscapes from mark-recapture data. Ecology 89, 542–554.

Pine, W.E., Pollock, K.H., Hightower, J.E., Kwake, T.J., Rice, J.A., 2011. A review of tagging methods for estimating fish population size and components of mortality. Fisheries 28, 10–23.

R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Schaub, M., Gimenez, O., Schmidt, B.R., Pradel, R., 2004. Estimating survival and temporary emigration in the multistate capture-recapture framework. Ecology 85, 2107–2113.

Schwarz, C.J., Schweigert, J.F., Arnason, A.N., 1993. Estimating migration rates using tag-recovery data. Biometrics 49, 177–193.

Sippel, T., Eveson, J.P., Galuardi, B., Lam, C., Hoyle, S., Maunder, M., Kleiber, P., Carvalho, F., Tsontos, V., Teo, S.L.H., Aires-da-Silva, A., 2015. Using movement data from electronic tags in fisheries stock assessment: a review of models, technology, and experimental design. Fish. Res. 163, 152–160.

Vermard, Y., Rivot, E., Mahevas, S., Marchal, P., Gascuel, D., 2010. Identifying fishing trip behavior and estimating fishing effort from VMS data using Bayesian Hidden Markov Models. Ecol. Modell. 221, 1757–1769.

Wetzel, C.R., Punt, A.E., 2011. Performance of a fisheries catch-at-age model (Stock Synthesis) in data-limited situations. Mar. Freshw. Res. 62, 927–936.

Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.

Yamamura, K., Kishita, M., Arakaki, N., Kawamura, F., Sadoyama, Y., 2003. Estimation of dispersal distance by mark-recapture experiments using traps: correction of bias caused by the artificial removal by traps. Popul. Ecol. 45, 149–155.