Title: **Exploring the utility of different tag-recovery experimental designs for use in spatially explicit, tag-integrated stock assessment models**

Authors: Daniel R. Goethel[1,a], Katelyn M. Bosley[b], Dana H. Hanselman[c], Aaron M. Berger[b], Jonathan J. Deroba[d], Brian J. Langseth[e], Amy M. Schueller[f]

Affiliations: [a] Sustainable Fisheries Division, Southeast Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 75 Virginia Beach Drive, Miami, FL 33133, USA.

[b] Fisheries Resource and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2032 S.E. OSU Drive, Newport, OR 97365, USA.

[c] Auke Bay Laboratories, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 17109 Point Lena Loop Road, Juneau, AK 99801, USA.

[d] Northeast Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 166 Water Street, Woods Hole, MA 02543, USA.

[e] Pacific Islands Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 1845 Wasp Blvd., Bldg. 176, Honolulu HI, 96818, USA.

[f] Beaufort Laboratory, Southeast Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 101 Pivers Island Road, Beaufort, NC 28516, USA.

Contact Author: [1] Daniel Goethel
75 Virginia Beach Dr.
Miami, FL 33133, USA
305-365-4490
daniel.goethel@noaa.gov

Abstract

The need for spatial stock assessment models that match the spatiotemporal management and biological structure of marine species is growing. Spatially explicit, tag-integrated models can emulate complex population structure, because they are able to estimate connectivity among population units by incorporating tag-recovery data directly into the combined objective function of the assessment. However, the limited scope of many small-scale tagging studies along with difficulty addressing major assumptions of tagging data has prevented more widespread utilization of tag-recovery data sets within tag-integrated models. A spatially explicit simulation-estimation framework that simulates metapopulation dynamics with two populations and time-varying connectivity was implemented for three life history (i.e., longevity) scenarios to explore the relative utility of tagging data for use in spatial assessment models across a range of tag release designs (e.g., annual, historical, periodic, and opportunistic tagging). Model scenarios also investigated the impacts of not accounting for incomplete tag mixing or assuming all fish were fully selected (i.e., that the age composition of tagged fish was unknown). Results demonstrated that periodic tagging (e.g., releasing tags every five years) may provide the best balance between tag program cost and parameter bias. For cost-effective tagging programs, tag releases should be spread over a longer time period instead of focusing on release events in consecutive years, while releasing tags in tandem with existing surveys could further improve the practicality of implementing tag-recovery experiments. However, care should be taken to fully address critical modeling assumptions (e.g., by estimating tag mixing parameters) before incorporating tagging data into an assessment model.

Highlights:

1. Including tagging data improved spatial assessments regardless of release design.
2. Periodic releases balanced tradeoffs between tag program cost and parameter bias.
3. Time-varying movement was estimable with informative periodic tagging data.
4. Violation of tagging assumptions increased parameter bias more than ignoring movement.
5. Estimating tag mixing parameters was feasible and eliminated associated bias.

Keywords: spatial models, tag-integrated models, stock assessment, connectivity, tag-recovery population structure, stock identification, tag mixing

1. Introduction

In recent years, advocacy for the development and implementation of spatial stock assessment models to support the often complex network of spatiotemporal fisheries management regulations has increased (Berger et al., 2017; Punt et al., 2019a,b). Spatially explicit models can directly account for spatial population structure and connectivity, while matching the scale at which data are collected and management actions enacted (Goethel et al., 2011; Berger et al., 2017; Rogers et al., 2017). However, the performance of spatial models depends on understanding the underlying spatial structure to ensure independent population units are being adequately identified and modeled (Kerr et al., 2016; Cadrin et al., 2019). As the scale of spatial assessment models becomes finer, it requires estimating a rapidly increasing number of additional parameters to account for connectivity, independent recruitment events, or biological parameters for each population unit modeled (Cope and Punt, 2011; Goethel et al., 2011; Punt, 2019b). To make estimation feasible, spatial assessments often utilize simplifying assumptions (e.g., functional forms for movement; Carruthers et al., 2015) or share parameters among population units, such as productivity (e.g., Punt et al., 2000) or selectivity (e.g., Thorson and Wetzel, 2016). Simulation testing has demonstrated that models which directly account for spatial structure often reduce bias compared with assuming no structure exists (i.e., panmictic assessments; Ying et al., 2011), implicitly modeling spatial structure (i.e., areas-as-fleets assessment approaches; Punt et al., 2015, 2016, 2017b, 2018), or ignoring movement among units (i.e., closed population models; Hulson et al., 2011; Goethel et al., 2015b;).

When explicitly incorporating spatial structure within an assessment model, it is often necessary to account for connectivity among population units, even though movement parameters may be poorly estimated and imprecise when no tagging data exist (Goethel et al., 2015b; McGilliard et al., 2015; Punt, 2018, 2019a). Parametrizing and identifying connectivity dynamics has become a focal issue for spatial assessment models, because misdiagnosing connectivity dynamics can result in a spatial model that performs as poorly as nonspatial assessments (Goethel et al., 2015b; Lee et al., 2017; Cadrin et al., 2019; Punt, 2019b). Early spatial assessment models relied on external estimates of movement typically from tagging analyses, which were then incorporated into the assessment as fixed parameters (e.g., Beverton and Holt, 1957; Quinn et al., 1990). As data quality and computing power have improved, connectivity rates have increasingly been treated as estimable parameters. By utilizing integrated assessment models (Maunder and Punt, 2013), preprocessed data from a variety of auxiliary sources can be incorporated in the assessment utilizing a combined objective function to estimate parameters. For instance, tag recaptures can be predicted in a sub-model using the same parameter values for both the tagged and untagged populations (e.g., Maunder, 1998). The combined likelihood approach of integrated models ensures consistency of assumptions and enhances estimates of uncertainty compared to the discrete two-step method of early spatial models (Maunder 1998, 2001). Additionally, by incorporating an additional data source (i.e., tagging data), tag-integrated assessment models utilize additional information to help estimate important parameters, such as fishing mortality, natural mortality, and, in spatially-explicit models, movement (Goethel et al., 2011; Punt, 2019b).

Implementing spatial tag-integrated models can be more demanding than nonspatial counterparts given the increased complexity of the modeling approach and resulting peer-review process (Berger et al., 2017), but assessments for a number of marine species have been improved through application of spatially explicit, tag-integrated models (e.g., Australian school shark, *Galeorhinus galeus*, Punt et al., 2000; South Pacific tunas using MULTIFAN-CL, Hampton and Fournier, 2001; and South African sardine, *Sardinops sagax*, de Moor et al., 2017). A number of simulation frameworks have explored the performance of spatial, tag-integrated models, particularly in comparison to spatial assessments that do not use tagging information (e.g., Maunder, 2001; Hulson et al., 2011, 2013; Goethel et al., 2015b; Vincent et al., 2017). Most studies have concluded that, when available, tagging data can greatly improve the performance of spatial assessment models by increasing the precision and accuracy of movement rates and reducing parameter confounding among recruitment and connectivity estimates (Hulson et al., 2011; Goethel et al., 2015b; Cadrin et al., 2019).

However, the spatiotemporal extent of tagging (or other auxiliary) data needed to reliably estimate complex movement patterns in spatial assessment models remains relatively unknown. Given resource limitations for fisheries data collection and assessment, identifying tradeoffs between modeling complex movement patterns and the extent of tagging data needed to inform movement parameter estimation is needed. A generalized spatially-explicit simulation-estimation framework was developed to determine the type of data (e.g., tag-recovery information) along with the complexity of movement parametrization required to reliably estimate population-specific parameters (e.g., biomass and fishing mortality trends) in spatial stock assessment models. The tradeoffs between the cost of various tagging program designs and resulting parameter bias in tag-integrated models were then identified. The framework involved simulating common fishery data and a tag-recovery study for a two population metapopulation connected through time-varying movement, then applying a variety of spatial assessment models to the simulated pseudo-data and comparing model performance. Simulation scenarios were placed into five groups to explore how 1) tagging time series, 2) tag deployment, 3) adherence to tagging data assumptions, 4) life history, and 5) movement parametrization impacted estimates from the applied assessment models. To address our objectives, we compared an estimation model that incorporated tagging data and estimated movement to ones that did not include tagging data or ignored movement. We also compared tag-integrated models that utilized perfectly implemented tagging studies to those utilizing tagging data where important assumptions of the tagging experimental design were violated (e.g., incomplete tag mixing occurred or the age of tagged fish was unknown). The results of the study provide new insight on the role of tagging data in implementing reliable spatial assessment models, the utility of different tag-recovery experimental designs for tag-integrated assessments, and the potential pitfalls of incorporating tagging data into assessments.

2. Methods

2.1 Overview

A simulation-estimation framework was developed, wherein common fisheries data (e.g., fishery catch and fishery-independent survey information including associated age compositions) and a tag-recovery study were simulated with measurement error. An assessment (estimation) model was then fit to the simulated 'observed' pseudo-data and estimates of parameters were compared to the true values used in the operating model. To explore the influence of the experimental design of a given tagging study along with model assumptions (i.e., of the tagging study or the assessment model spatial structure) on estimation model performance, a total of 16 scenarios were carried out (with an additional 56 scenarios provided in the supplementary material). Scenarios were placed in five groups (i.e., tagging time series, tag deployment protocols, tag data assumptions, life history, and movement parametrization). Scenario names are provided in italics (and used throughout the text) with full details of the main model runs provided in detail in section 2.4 (*Simulation Scenarios*).

The operating model was implemented to simulate the dynamics of a metapopulation (as defined in Goethel and Berger, 2017) consisting of two interconnected populations with differing demographics and productivity regimes. Reproductive mixing occurred among populations through the movement of mature individuals, but each population was assumed to maintain its own larval pool and stock-recruit function. Instantaneous box-transfer movement was assumed at the beginning of the year and once fish moved into another area they assumed the reproductive dynamics and demographics of the population residing in that area, which implied that environment was the main driver of life history (not genetics). Population dynamics were simulated for thirty years starting from an input initial abundance-at-age and applying random annual deviations for recruitment, fishing mortality, and movement to encapsulate variation. Pseudo-data were generated for each year of the model with measurement error simulated for each data source using stochastic processes based on an assumed underlying probability distribution. For each scenario, a total of 500 runs were simulated, and, for each run, the data set differed due to the realized measurement error. Each run maintained the same population dynamics (i.e., random deviations on population parameters were constant) and differed only in the implemented measurement error. Similarly, across all scenarios, associated run numbers were identical in terms of both population random deviations and realized measurement error (i.e., across all scenarios, run number one had identical population trajectories and data sets) to facilitate comparison across simulation scenarios.

Spatially-explicit stock assessments were applied to the various simulated, thirty-year time series of pseudo-data (with or without fitting tag recaptures). The assessment models matched the operating model dynamics except for the parametrization of movement, which varied from ignoring movement to estimating annual rates. Error, precision, and stability were assessed for each scenario based on model performance across all converged runs.

The operating model was described in Goethel and Berger (2017, using the metapopulation configuration) with the addition of simulated tag-recovery pseudo-data. The estimation models were generalized versions of those outlined in Goethel et al. (2011) and implemented in Goethel et al. (2015a,b) with further refinements, particularly in the handling of tagging data. Both models were coded in AD Model Builder (Fournier et al., 2012) and can be downloaded from the Github repository (https://github.com/dgoethel/tag-integrated-model).

2.2 Operating model

The two population, metapopulation operating model was parametrized to simulate the dynamics of a relatively short-lived (plus group at eight years), fast growing species. Each population maintained typical assumptions for species of medium longevity including moderate levels of natural mortality (*M*, instantaneous value of 0.2 and 0.25 for population one and population two, respectively), interannual variation in recruitment (σ_R, value of 0.5 and 0.55 for population one and population two, respectively), connectivity among populations (*T*, maximum annual movement rate of 20% and 25% of the population for population one and population two, respectively), and fishing mortality (that assumed a dome-shaped time trajectory). Simulations were not meant to mimic the dynamics of any specific species, but were set up to resemble general biological dynamics that may apply to several species groups (e.g., certain coastal pelagic species, tunas, ground fish, or reef fish species). Variation in parameters (along with stock-recruit relationships) among populations helped emulate metapopulation dynamics, because population units often demonstrate unique demographic and reproductive rates in metapopulation systems (see Goethel and Berger, 2017). The sequential order of events in the operating model involved: (1) spawning; (2) recruitment to the population and fishery; (3) release of tagged fish, if tagging takes place in that year; (4) instantaneous movement of tagged and untagged fish among populations; and (5) continuous natural mortality and removals due to harvest throughout the year, including tag recaptures with reporting rates of 70% and 80% (for population one and population two, respectively). For a complete description of the population dynamics see Supplementary Material SM.1 (including Table SM1-2 and Figures SM 1-2 for operating model input parameters, as well as Goethel and Berger, 2017, including Figure 2 therein for a schematic illustrating the population dynamics).

2.2.1 Data generation

The operating model produced five population-specific sets of pseudo-data: (1) age compositions from the catch; (2) fishery-independent survey age compositions; (3) total yield; (4) fishery-independent survey biomass; and (5) tag recaptures. Measurement error was incorporated into each data set based on an underlying error assumption (i.e., lognormal error for fishery yield and survey biomass along with multinomial error for fishery and survey age compositions and tag recapture states; Table 1). For a full description of the pseudo-data generation process see Supplementary Material Section SM1.3 on the incorporation of measurement error.

Differences in tagging experimental design were the primary way in which operating models differed, particularly in how tags were released across years, populations, and ages. A multiyear Brownie tagging model (Brownie et al., 1993) imbedded directly within the operating model simulated the tag-recovery pseudo-data across multiple release and recapture events (following the estimation model equations of Lauretta and Goethel, 2017). In each year of the simulation, a new tag cohort could be released into the population, where a cohort was defined by the combination of year, age, and population of release. The tag release protocol was defined by a combination of four independent processes: the number of tags released, the frequency of tag release events, the population distribution of tags, and the age distribution of tags. The sequential order of tagging dynamics involved: (1) a simulated release event at the beginning of the year that defined the number of fish released in a given cohort; (2) instantaneous movement post-tagging, with potential for incomplete mixing of the tagged and untagged population in the year of release (i.e., different movement rates for tagged fish); (3) continuous mortality throughout the year (with potential for incomplete mixing causing different fishing mortality in the year of release), which resulted in recaptured tags that were tallied by cohort and population of recapture (and accounted for non-reporting of tags); (4) repetition of this sequence in the following year starting at step (2) for tagged fish that survived, which continued until a mortality event or the maximum life of the tag was reached (see Supplementary Material section SM.1.3 for a full description of the tag dynamics).

There were two types of tag release designs in the model: fixed and opportunistic. A majority of scenarios utilized a fixed design where a set number of tags were released during each release event, which occurred in pre-determined years and populations throughout the time series. Opportunistic tagging designs utilized probability distributions to determine whether a tag event occurred in a given year (Bernoulli distribution, $p = 0.7$) or population (Bernoulli distribution, $p = 0.6$) and were also used to set the number of tag releases in a given release event (uniform distribution; see Table SM2 for the inputs assumed for each tagging distribution). The opportunistic tagging scenarios were meant to emulate, for example, multiple patchwork studies over time (e.g., a handful of independent, short-term studies). Although the simulations do not account for other potential issues with these types of tagging programs (e.g., tagging only certain age or size classes), they provide insight to the usefulness of patchwork tagging programs.

For the fixed tagging designs, a total of 5,000 tags were released during each release event. Tags were assigned to a release cohort by apportioning the total releases to a population based on the relative survey biomass and distributing across ages within a population relative to survey age compositions in the given population (see Table 2 for the details of the *Base* scenario tagging inputs). The tag deployment dynamics were parameterized so that the number of tags was much less than 1% of initial population abundance and that fish were tagged using the same gear as the survey resulting in the same age distribution. The age of tagged fish was thus provided to the assessment model without error. Although these assumptions are reasonable for carefully designed tagging studies, the known age of release assumption would be more difficult to adhere to in real-world situations. Therefore, a sensitivity run was explored that assumed the age of tagged fish was unknown (see section 2.4, *Simulation Scenarios*).

Movement was assumed to occur immediately following tagging, which resulted in tags being available for recapture from each cohort in each population in the release year. However, in the year of release, the model was able to account for incomplete mixing of tagged fish and untagged fish by scaling movement and fishing mortality by associated proportionality coefficients (see the tag data assumptions scenarios, Table 3). Tag recaptures by cohort in a given year and population were calculated using Baranov's catch equation assuming a continuous year-long process of mortality and harvest and discounting tags for non-reporting based on a reporting rate parameter. It was assumed that each tag had a lifespan of five years (after which, if a tagged fish was still alive, it was placed in the not recaptured state for that cohort), and there was no tag loss or tag induced mortality. The basic tagging dynamics were implemented in all scenarios unless otherwise noted in section 2.4 (Table 3).

2.2 Estimation models

The estimation models matched the operating model parameterization (including natural mortality and reporting rates being fixed at the true values), with the exception of movement (estimated in two year time blocks). Each estimation model was implemented using an integrated statistical catch-at-age framework (Maunder and Punt, 2013) based on a generalized version of the assessments used in Goethel et al. (2015a,b; see section SM.2 in the Supplementary Material for a complete description of the estimation model). The variance terms and effective sample size (*ESS*) for each likelihood component were also taken directly from the operating model (Table 1), because error misspecification was not considered here. Variants of the estimation model included: (a) the *Base* scenario model which matched the operating model except that movement was estimated in two year time blocks; (b) a spatial model which matched the *Base* scenario, but did not incorporate tagging pseudo-data (*No_Tag*); (c) a closed population model that treated each population as independent units assuming no movement between them (*No_Move*); (d) the *Base* scenario model, but with parameters estimated to account for incomplete tag mixing (*Est_Tag_Mx*); (e) the *Base* scenario model, but assuming the age of tagged fish was unknown forcing the estimation model to fit age-aggregated tagging cohorts (*No_Age_Tag*; see Table 3 for a summary of scenarios).

2.3 Evaluation of model performance

The performance of each estimation model scenario was compared based on bias and precision in estimates of population parameters (e.g., recruitment, fishing mortality, biomass, and movement rates). Mean relative error (MRE; an overall measure of bias) and the median absolute relative error (MARE; a measure of bias and variability) for a given model parameter were calculated by population aggregated across the time series (i.e., calculated using the thirty years of estimates across all 500 model runs within each scenario). Model stability, an indicator of over-parametrization and robustness, was addressed by calculating the proportion of runs that an estimation model converged.

2.4 Simulation scenarios

Model scenarios were placed in five groups, which included tagging time series length, tag deployment protocols, tag data assumptions, life history, and movement parametrization. Scenario names are provided in italics (and used throughout the remaining text) with full details of the main model runs provided in Table 3. Additional sensitivity runs are summarized in the Supplementary Material (Table SM3).

The setup of the *Base* simulation scenario tag release design was meant to balance the relative cost of the tagging program (i.e., releasing tags every five years) with parameter estimation performance, particularly for movement parameters, to demonstrate a cost-effective model of intermediate complexity. The parametrization of movement in the estimation model balanced model complexity against precision of parameter estimates by estimating movement in two-year time blocks (as was suggested by Goethel et al., 2015b for estimation of time-varying movement in spatial assessment models) instead of annually. Each of the scenario runs was compared to the *Base* model scenario results to explore how changes in the tagging program or alternate assumptions impacted estimation model performance.

Group 1: tagging time series

There has been limited exploration of alternate tag release designs to determine whether the frequency and timing (relative to the overall assessment time series) of release events may be more important factors than overall length of a tagging time series. Several common short-term tag release designs (e.g., releases over five consecutive years) were simulated and differed according to the point in the time series at which they were implemented [e.g., beginning (*Tag_Beg_5*), middle (*Tag_Mid_5*), and end (*Tag_End_5*) of the time series]. An annual tagging time series where tags were released every year (*Tag_Yrly*) was also implemented. These were compared with more unique designs that allowed for periodic tagging, which were spread out across the entire time series [e.g., every five years (*Base*) and every ten years (*Tag_Evy_10*)]. A spatial model that did not incorporate tagging was also implemented (*No_Tag*).

Group 2: tag deployment

Scenarios also included different design aspects for how tags were released including how tag releases were distributed across populations [e.g., proportional to survey biomass by population (*Base*) or releasing tags in only one population (*Tag_Area_2*)]. A fully opportunistic tagging design was also implemented (*Opp_Tag*) wherein the number of tags released was defined by a uniform distribution, the probability of a tag release event in a given year was determined by a Bernoulli distribution (with potential release event years matching the *Base* scenario), and the probability of a release event occurring in a given population was defined by an independent Bernoulli distribution (see Table SM2). This release design was meant to emulate a patchwork tagging program that released tags as funding became available or as a series of pilot projects over time with limited spatial scale.

Group 3: tag data assumptions

Two main assumptions of tag-recovery data, complete mixing of tags and known age structure of tags, were explored to determine how tag-integrated models performed when these assumptions were violated. To emulate incomplete mixing of tagged fish during the year of release, simulations were implemented wherein tagged fish were assumed to have a much higher residency (i.e., randomly distributed around an average residency rate of 90%) and lower levels of fishing mortality (i.e., 50% of the associated fishing mortality on untagged fish). Associated estimation models then either ignored tag mixing (*No_Tag_Mx*) or estimated independent parameters for movement and fishing mortality for tagged fish in tag release years (*Est_Tag_Mx*). For the estimation model that accounted for incomplete tag mixing, cohort-specific fishing mortality and movement parameters were estimated directly for tagged fish in the year of release.

The *Base* model scenario assumed that the age composition of all tagged fish in a cohort was known (e.g., by either taking non-invasive scale samples to determine age directly or applying age-length keys to the length composition of tagged fish); however, directly aging tagged fish is often not feasible, and age-length keys may result in biased age composition information. Therefore, to provide an indication of the maximum bias that might be expected when the age structure of tagged fish was unknown, the *No_Age_Tag* scenario simulated age-based tagging dynamics with the associated estimation model ignoring age structure in the tagging sub-model. For the estimation model, the input tag releases were summed across ages, and the model then calculated predicted tag-recaptures assuming 100% selectivity and with age (i.e., the age subscript) removed from the calculations. In the objective function, the tag-recapture pseudo-data were summed across ages, and the pooled pseudo-data was fit to the tag-recaptures predicted by the assessment model. The inherent process error due to age-based tagging dynamics in the operating model that was not accounted for in the estimation model provided a simple approximation to the error that might result from unknown ages of tagged fish.

Group 4: life history

To enable moderate generalization of the findings beyond the single life history utilized for all other scenarios, both long-lived (*LL_Evy_5*) and short-lived (*SL_Evy_5*) life history scenarios were implemented. The long-lived scenario doubled the number of ages to sixteen as well as doubling both the age at 50% maturity and selectivity and halving the natural mortality to 0.1. On the other hand, the short-lived scenario halved the number of ages to four along with halving the age at 50% maturity and selectivity, whereas natural mortality was doubled. Both life history scenarios assumed the same tagging dynamics as the *Base* scenario (i.e., releasing tags every five years). Although the life history scenarios were rudimentary approximations of either fast growing small pelagics (i.e., the short-lived scenario) or relatively slow growing ground fish or deep-water species (i.e., the long-lived scenario), they provided an indication of the robustness of the *Base* scenario tagging methodology across a variety of life history types.

Group 5: movement parametrization

Several alternate movement parametrizations were implemented to illustrate how ignoring movement (*No_Move*) or assuming constant movement rates (*Cnst_Move*) could potentially bias resulting parameter estimates compared to estimating movement in two-year time blocks (*Base*). Other exploratory scenarios were included in the supplementary material (see Table SM3) that compared how different movement parameter time blocks performed [estimating yearly movement (*Move_Yrly*), estimating yearly movement with yearly tag releases (*Yr_T_Tag_Yr*), and estimating movement in five year time blocks (*T_Blk_5_Yr*)].

2.5 Comparison of relative tag program cost

The relative cost of each tagging experimental design was calculated as an approximation of actual tagging program costs based on design features (i.e., the number of tags per year, number of populations in which tagging occurred, and number of years of tag releases). Cost for each tagging scenario was determined relative to the *Base* scenario tagging program [i.e., 5,000 tags released every five years (for a total of seven years of releases) across two populations] where each tagging design component (i.e., population, year, and every 5,000 tags released) was assigned a unit cost of one. Therefore, the *Base* tagging scenario (and both life history scenarios) had a total cost of 14 units (two populations*seven years*one unit of tags). All other tagging programs were scaled up or down based on the relative number of populations and years in which tagging occurred. The cost of the opportunistic tagging scenario was discounted by 25%, because this scenario was meant to represent tagging programs that operated as opportunity arose (implying a lower cost). Expenses related to tag recoveries (e.g., advertising and tag rewards) were assumed to be similar across tagging designs, and these costs were not included. Plots were then developed to illustrate relative tag program cost and resulting MARE values across tag release scenarios, which allowed comparison of the cost of a tagging program versus the expected improvement in tag-integrated model performance.

3. Results

Base scenario performance

The *Base* model scenario was first fit to the simulated pseudo-data without measurement error as both a self-consistency run and as a basis of comparison to demonstrate the impact of measurement error on model estimation. When fit to pseudo-data without measurement error, the *Base* scenario was able to replicate the population-specific biomass trends almost exactly (Figure SM2). Because movement was estimated in two-year time blocks, the trend tended to follow the mean level of movement across the two years for which each movement parameter was estimated. Although the pattern reflected the true movement dynamics relatively well, the estimation model was not able to match the exact values in any given year due to the inherent mismatch in the operating model and estimation model parameterizations. However, the two-year time block parametrization of movement performed much better than yearly movement estimation, because the latter was over-parametrized (Figure SM2).

When fit to pseudo-data with measurement error, the *Base* scenario also performed well, but with lower precision in estimates (Tables 4-5, Figures 1-2). Biomass estimates over the time series were unbiased (MRE near zero; Table 4) with high precision (MARE ranged from 1.47 to 4.63; Table 5, Figure 1). Estimation of fishing mortality in both populations demonstrated slight overestimation (MRE ranged from 1.48 to 7.32; Table 4), but high precision (MARE was between 6.09 and 7.32; Table 5, Figure 1). Population specific recruitment estimates tended to be slightly overestimated (MRE between 1.86 and 7.16; Table 4) with moderate imprecision (MARE ranging from 9.37 to 12.94; Table 5, Figure 1). System-wide estimates of both recruitment and biomass tended to be much more accurate and precise than did population-specific estimates. Movement parameters were the most biased (MRE between 5.49 and 6.71; Table 4) and imprecise (MARE between 22 and 25.5; Table 5, Figure 1). Terminal year parameter estimates demonstrated higher levels of bias, particularly in population-specific recruitment estimates where population one recruitment tended to be overestimated and vice versa for population two (Figure 2).

The *Base* scenario demonstrated limited parameter correlation resulting in high model stability. Some minor correlations occurred among recruitment parameters and among initial abundance parameters, which was to be expected given the relative lack of information in the data to support independent estimation of many of these parameters. However, these correlations did not influence model stability. The overall convergence rate of the *Base* scenario was 98% (Table 3). High convergence was common across all simulation scenarios indicating that there were no major issues stemming from parameter correlation or general model instability. However, the short-lived life history (*SL_Tag_Evy_5*) scenario had a convergence rate of 89%, which was reflective of the difficulty it had in estimating movement parameters.

The results of the alternate scenarios relative to the *Base* scenario are discussed by scenario group with emphasis placed on the more novel findings. Results from scenarios not discussed in the main text can be found in the supplementary material, because these scenarios did not add significantly to the primary findings or simply supported conclusions from previous studies (Figures SM3-SM10).

Group 1: tagging time series

The model without tagging data (*No_Tag*) demonstrated high imprecision in parameter estimates, most noticeably in movement rates (MARE of 72 to 91; Table 5, Figures 1-2). Similarly, levels of bias for fishing mortality in population one increased (MRE of 14.6; Table 4) compared to the *Base* scenario. However, estimates of biomass were relatively unbiased (MRE ranged from -1.36 to 2.96), albeit with higher imprecision than the Base scenario (MARE ranged from 1.88 to 8.57; Tables 4-5). Although the no tagging model did not have convergence issues, there was strong correlation between and among movement and recruitment parameters that caused some runs to estimate zero recruitment in an area with a correspondingly inflated movement of fish into that area (i.e., all recruitment was in one population with high emigration from that population to allow those recruits to then inhabit the other population; Figure 3). Tagging more frequently (i.e., the *Tag_Yrly* scenario) slightly reduced bias and imprecision, whereas tagging less frequently (*Tag_Evy_10*) had the converse effect, although neither scenario demonstrated patterns that differed greatly from the *Base* scenario. Short-term, clumped tagging programs (i.e., *Tag_Beg_5*, *Tag_Mid_5*, *and Tag_End_5*) all performed similarly with generally elevated bias and imprecision compared to the *Base* scenario (Tables 4-5, Figures 1-2). Tagging at the end of the time series resulted in higher parameter bias across the time series (e.g., in population one fishing mortality; Figure 1), yet better terminal year estimates of fishing mortality and movement (Figure 2).

Group 2: tag deployment

Tagging opportunistically (*Opp_Tag*) led to similar performance as the *Base* scenario, but with increased levels of bias and imprecision in terminal year estimates (Figure 2) and movement parameter values (MARE between 44 and 47; Tables 4-5; Figure 1). Tagging only in population two (*Tag_Area_2*) performed similarly to the *Opp_Tag* scenario, but with improved movement estimates (even compared to the *Base* scenario; MRE ranged from -3.07 to 1.67) and increased bias in population two recruitment (MRE = 18.19; Tables 4-5, Figures 1-2).

The impact of tagging data and associated tag release design was most clearly demonstrated by looking at the time series of movement estimates, recruitment, and fishing mortality (Figure 3). Without tagging data (*No_Tag*), the model was not able to accurately estimate movement rates, which led to a number of runs estimating zero recruitment in a given area, whereas the reduced information on mortality rates caused by not having tagging data led to higher imprecision in fishing mortality. The addition of tagging data (e.g., the *Base* scenario) immediately improved movement estimates starting in the first year of release and extended for the assumed lifespan of tags (i.e., five years) with decreasing impacts as fewer tags remained in the system. The immediate effect was most clearly seen for the *Tag_Mid_5* and *Opp_Tag* scenarios wherein movement parameters were highly imprecise until a release event occurred, while the precision slowly decreased following a release event (Figure 3). Similarly, precision and accuracy of both recruitment and fishing mortality were improved in years immediately following a release event. The periodic release design (i.e., releasing tags every five years) of the *Base* scenario allowed moderately precise movement parameters estimates, while providing high accuracy and precision of other model parameters over the entire time series (Figure 3). Although the annual tagging model (*Tag_Yrly*) greatly increased the precision of the movement parameters, the overall improvement in other median parameter estimates was minimal compared to the *Base* scenario.

Group 3: tag assumptions

Violation of the tag model assumptions led to the worst performing models in this study. For the model in which tag age was unknown (*No_Age_Tag*), bias levels were high with fishing mortality being underestimated (MRE between -1 and -15), which caused biomass estimates to be overestimated (MRE ranged from 3.5 to 14) and led to increased imprecision compared to the *Base* scenario (Tables 4-5, Figures 1-2). Not accounting for incomplete mixing when it was taking place (*No_Tag_Mx*) led to similar, but less extreme patterns in parameter bias and precision as the *No_Age_Tag* scenario (population specific biomass MRE was between 9 and 13 with fishing mortality MRE ranging from -6 to -14; Tables 4-5, Figures 1-2). The *Est_Tag_Mx* model was able to accurately estimate the scalars on fishing mortality (F_{MIX}) and the new movement rates for tagged fish in each year of release, which resulted in comparable parameter bias to the *Base* scenario with only moderately increased imprecision (e.g., movement rate MARE around 31; Tables 4-5, Figures 1-2). Ignoring incomplete mixing (i.e., the *No_Tag_Mx* scenario) caused severe underestimates of fishing mortality in release years leading to overestimation of biomass (Figure SM3). Conversely, when the model was allowed to estimate the scalar on fishing mortality (i.e., the *Est_Tag_Mx* scenario) to account for incomplete mixing, the bias was removed (Figure SM3).

Group 4: life history

Both the short-lived (*SL_Tag_Evy_5*) and long-lived (*LL_Tag_Evy_5*) life history scenarios performed similarly to the *Base* scenario (Figures 1-2). Although the short-lived scenario actually demonstrated lower bias compared to the *Base* scenario for some parameters (e.g., MRE in fishing mortality ranged from -1.77 to 2.2; Tables 4-5), it was unable to accurately estimate movement rates demonstrating higher bias and imprecision (MRE ranged from 18.74 to 32.51 and MARE ranged from 30.81 to 33.17; Tables 4-5). The long-lived scenario had slightly increased bias compared to the short-lived scenario, but precision was generally higher, particularly in estimates of movement rates (MARE ranged from 18.83 to 22.88; Table 4).

Group 5: movement parametrization

Ignoring movement (*No_Move*) was detrimental to model performance leading to inaccurate estimates of important parameters, including population-specific biomass (MRE ranging from -6.5 for population one to 13.14 for population two; Table 4), particularly in the terminal year (Figure 2); however, system-wide values tended to be relatively well estimated (e.g., biomass MRE = 1.43 and recruitment MRE = 1.73; Tables 4-5, Figures 1-2). The constant movement scenario (*Cnst_Move*) performed well with only slight increases in bias and imprecision compared to the *Base* scenario (Tables 4-5, Figures 1-2).

Comparison of relative tag program cost

Tagging every five years (i.e., the *Base* scenario) provided an adequate balance between a relatively inexpensive tagging program (compared to annual tagging, *Tag_Yrly*) and low resulting MARE for many population parameters compared to less resource intensive tagging programs with fewer release events [e.g., tagging every ten years (*Tag_Evy_10*), tagging in only one area (*Tag_Area_2*), or opportunistic tagging (*Opp_Tag*); Figure 4]. However, less intensive and easier to implement (and maintain) tag designs, such as opportunistic tagging (*Opp_Tag*), resulted in only a moderate increase in MARE with considerable cost savings.

4. Discussion

Modeling complex spatial dynamics in stock assessment models likely requires some form of auxiliary information, such as tag-recovery data, to inform connectivity and adequately estimate population trajectories. Previous spatially explicit tag-integrated simulation studies have focused on tagging data quality and quantity (e.g., Hulson et al., 2011, 2013; Goethel et al., 2015b; Vincent et al., 2017), but our results indicate that the frequency and distribution of tag releases over time and space may be as important for achieving accurate and precise parameter estimates. Longer time series of data inputs for an assessment, particularly tagging data, usually results in improved model performance (Goethel et al., 2015b). However, in the case of collecting tagging data, there are other factors (e.g., funding, weather, or availability of boat time) that may limit the ability to release and recapture tagged fish every year and at all locations. Most tagging studies do not match the spatial extent of the population or the longevity of the species, because they are typically financed by short-term grants. Given these common circumstances, our results provide an exploration of tradeoffs among tagging design cost and the expected benefits in terms of tag-integrated assessment model performance (Figure 4).

Our simulation scenarios were limited in their exploration of process error and spatiotemporal complexity (including the form of underlying movement dynamics) resulting in uncertainty estimates that are likely to be severely underestimated when compared to real world applications of spatial assessment models (e.g., when connectivity and tagging occur across entire ocean basins). Despite these caveats, there were a number of general results that are likely to be useful in future applications of tag-integrated assessments. For instance, when tag releases were spread across the assessment time series, the information content in tag recaptures improved parameter estimates for the entire length of the assessment period. Tag releases were not required every year, though, given that the *Base* model scenario, in which tagging occurred every five years, demonstrated similar performance to more frequent tag release scenarios (e.g., annual tag releases, *Tag_Yrly*). Performing periodic release events provides a tag recapture time series of sufficient length to improve assessment outputs at a substantial cost savings over annual tagging studies. These results also held across multiple life history types (e.g., short-, medium-, and long-lived species) indicating some degree of generalization was possible.

Releasing tags opportunistically across both years and populations (*Opp_Tag*) provided accurate parameter estimates at a substantially reduced cost of the tagging program compared to traditional fixed tag release designs (due to releases not occurring in every population and potential release year). Although any tag study must still adhere to the major assumptions for utilizing tagging data, these results indicate that tagging studies of limited scope (e.g., pilot studies or opportunistic tagging as funding becomes available) could still provide useful data for tag-integrated models. Similarly, tagging in only a single spatial unit (e.g., the *Tag_Area_2* scenario) can also be informative. However, when there are spatial tag deployment limitations it may be better to tag in the smaller, less productive population unit (see results for the *Tag_Area_2* scenario compared with those from the *Tag_Area_1* scenario in the Supplementary Material). By doing so, a stronger signal is provided regarding the emigration rates from and fishing mortality on the less productive population. Information on the population trajectories of less productive population units are important for spatial models, because signals in other data sources (e.g., landings and age composition) are often inundated by the larger population components (Goethel et al., 2015b; Vincent et al., 2017).

Short-term tagging studies (e.g., one time or clumped release events) provide bursts of information to the assessment that help stabilize the model by reducing correlation among movement and recruitment parameters (Goethel et al., 2015b; Cadrin et al., 2019). However, results indicated that a better use of funding for tagging programs would be to spread release events over a longer time period instead of implementing a limited number of release events in consecutive years. For instance, the main reason that the opportunistic tagging study performed well was because tag releases occurred across the time series, thereby providing information from multiple periods compared to the brief, single period snapshots provided by short-term studies. Given that many tag programs are funded by short-term grants, it may be difficult to optimize release designs in this way. Ideally, complimenting survey data by conducting intermittent tag release programs as part of existing survey designs (e.g., as is done with Alaskan sablefish; Hanselman et al., 2015) may produce the highest return on investment for funding agencies and would provide much needed information on movement that surveys alone often cannot provide. Identifying alternate data sources that can inform connectivity and be collected as part of survey protocols (e.g., natural tags, genetic information, or otoliths), as was done for South African sardine using parasite infestation rates (de Moor et al., 2017), provides a cost-effective alternative to implementing tagging programs. However, there may be unaccounted for costs (e.g., advertising) or impediments (e.g., time-varying reporting rate) to maintaining a longer time series of recaptures, which were not addressed in this study and would need to be considered for long-term periodic tagging programs.

Lack of tagging data (i.e., the *No_Tag* scenario) degraded performance compared to most of the models that included some form of tagging information. However, population-specific parameter estimates were still relatively unbiased. The main detriment was increased imprecision, which corroborates earlier studies comparing tag-integrated and spatial models without tagging data (e.g., Hulson et al., 2011; Goethel et al., 2015b). As discussed in depth in Goethel et al. (2015b) and Cadrin et al. (2019), the primary issue with spatial models that lack tagging data is that recruitment and movement parameters often become highly correlated. Although spatial models without tagging information often outperform similar models that assume no movement (as was the case when comparing the *No_Move* and *No_Tag* scenarios; Goethel et al., 2015b; McGilliard et al., 2015; Punt, 2019a), results often depend on the existence of high quality age composition data to inform movement parameter estimation in the spatial models. When age composition data are of poor quality (e.g., the *No_Tag_LQ* scenario provided in the supplementary material), estimating the parameters of a spatial model without tagging data may be more detrimental than ignoring movement, because there is increased probability of high estimation bias and model instability (e.g., high parameter correlation leading to unrealistic outcomes). The benefit of age composition data could also be seen in the life history runs where the short-lived life history scenario had a more difficult time estimating movement rates compared to the medium- (i.e., *Base*) and long-lived scenarios. These estimation difficulties are believed to be partly due to the relative lack of information contained in the condensed (i.e., fewer age classes) age compositions available for short-lived species, but was also influenced by each cohort only experiencing on average one tag release event (i.e., the average life span was four years, whereas the tag frequency was every five years).

Mis-specifying movement parametrization (e.g., assuming constant movement when it is actually time-varying) can be as detrimental as ignoring movement altogether or implicitly accounting for spatial dynamics through areas-as-fleets models (Hulson et al., 2013; Goethel et al., 2015b; Lee et al., 2017; Li et al., 2018). The constant movement (*Cnst_Move*) scenario in the current study performed moderately well, albeit with strong cyclical bias in biomass. Because there was not a strong trend over time in movement rates in the operating model, the constant movement model was not penalized for its inability to estimate annual deviations in the movement rates. Additionally, ignoring movement (e.g., the *No_Move* scenario) may lead to reasonable estimates of total biomass, which suggests that panmictic assessments could also provide adequate domain scale estimates (e.g., Li et al., 2015). However, the situations for which individual population dynamics and connectivity could be ignored are likely to be limited given the potential for depletion of population components (Ying et al., 2011; Goethel et al., 2011; Guan et al., 2013; Kerr et al., 2016; Punt et al., 2018).

Given the relatively simple simulated movement dynamics (i.e., time-varying without trend) compared to the often complex ontogenetic patterns observed in real-world applications, the results of this study are likely to be overly optimistic. For instance, if age-based movement occurs, it is likely that estimating movement for long-lived species will be much more difficult given the greatly increased number of movement parameters that would need to be estimated. Therefore, future research should further investigate the feasibility of estimating more complex movement dynamics with limited or no tagging information along with the associated bias from ignoring age- and time-varying movement, given that connectivity dynamics are unlikely to be static across either time or age.

The benefits of including tagging data in an assessment must be weighed against the increase in the number of parameters to be estimated and the potential for violation of critical tagging assumptions. For instance, in the scenarios where the age of tagged fish was unknown (*No_Age_Tag*) or incomplete mixing was ignored (*No_Tag_Mx*), incorporating tagging data led to biased models that often performed worse than not including any tagging data. These results are important because many tagging studies do not have exact age at release information, and homogenous mixing of tagged individuals across large-scale spatial domains is effectively impossible. Additionally, it is difficult to fully verify or fulfill all of the assumptions of tag-recovery data (e.g., that the dynamics of the tagged fish are representative of the general population or that the age composition of tagged fish and the untagged population overlap appropriately; Ziegler, 2013). More research is also needed on best practices for incorporating tagging data, particularly with regard to data weighting (Punt, 2017a) and accounting for over-dispersion caused by non-independence of sampled (tagged) fish (i.e., using alternate likelihood functions; Hanselman et al., 2015).

However, many tagging data assumptions can be directly accounted for by adjusting the parametrization of tagging models. For instance, it is common practice for tagging models to estimate tag mixing parameters (as was done in the *Est_Tag_Mx* scenario; e.g., Hoenig et al., 1998; Hampton and Fournier, 2001; Waterhouse and Hoenig, 2011). External analyses can also be performed to address tag mixing assumptions (e.g., Kolody and Hoyle, 2015) and tag recaptures that are deemed to have been at-large for too short a time period to undergo full mixing with non-tagged fish can be removed (e.g., Punt et al., 2000). Similarly, the bias associated with the *No_Age_Tag* scenario is likely to be extreme, because information on the age composition of tagged fish can often be derived by taking scale samples of all tagged fish or by collecting otoliths of recaptured fish. The lengths of tagged fish can then be assigned to an age class using age-length keys [as is done in MULTIFAN-CL (Hampton and Fournier, 2001) and other applied tag-integrated models (e.g., Cadigan, 2016; ICES, 2017)], thereby avoiding the full selection assumption of the *No_Age_Tag* scenario. Similarly, length data of tagged fish can be fit directly without converting to age composition, but more work is needed to explore the performance of tag-integrated models using only length data from tagged fish. In most cases, the benefit gained from incorporating tagging data will outweigh potential pitfalls as long as the critical assumptions are carefully considered and tag-integrated models are parametrized accordingly

There are considerable opportunities for incorporating underutilized spatially-explicit data sets, which exist in many fisheries agencies (e.g., tag-recovery data, electronic tagging, genetics, otolith microchemistry, or vessel monitoring system landings data), into integrated assessment frameworks to inform complex spatial dynamics (Goethel et al., 2011; Berger et al., 2017; Li et al., 2018). By matching the flexibility of integrated analysis with fine-scale spatial models, data can be used at the scale at which they were collected (e.g., by implementing distribution models within the assessment framework; Berger et al., 2017). A good deal remains to be learned about the implementation and parametrization of spatial assessment models, yet tag-integrated frameworks are clearly an informative option for representing complex real-world spatial dynamics. The utility of spatial models depends on the goals of management and the importance of understanding fine-scale dynamics for a given species (Berger et al., 2017; Punt et al, 2019a,b). Continued work is needed to identify robust management strategies when complex spatiotemporal dynamics exist (e.g., Punt et al., 2017b). Despite the simplicity of the simulation framework we applied (i.e., limited spatiotemporal complexity and process error, known natural mortality and reporting rates, and perfect alignment of biological and assessment units), our results provide insights into the importance of accounting for spatial population structure in assessment models and the role that tagging data, even when collected at limited spatiotemporal scales, can have for informing connectivity patterns and improving population-specific parameter estimates from spatially-explicit models.

5. Acknowledgements

The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect those of NOAA or the Department of Commerce. We thank Mark Maunder and the Center for the Advancement of Population Assessment Methodology (CAPAM) for organizing the workshop on “Development of spatial stock assessment models” (La Jolla, CA, 2018) and this special issue. The manuscript was greatly enhanced by the insightful reviews of Matt Lauretta, two anonymous reviewers, and the editor, Steve Cadrin. Model development was aided by discussions with Terry Quinn, Kari Fenske, Lisa Kerr, and a variety of participants at the CAPAM spatial workshop. Funding for this project was provided by NOAA’s National Marine Fisheries Service Office of Science and Technology branch under the Stock Assessment Analytical Methods (SAAM) grant and the National Research Council Research Associate Program (NRC-RAP). D. Goethel would like to note his personal gratitude to Terry Quinn for the support and inspiration provided throughout his career, especially for pursuing spatial research topics; he will be deeply missed.

6. References

Berger, A.M., Goethel, D.R., Lynch, P.D., Quinn II, T., Mormede, S., McKenzie, J., and Dunn, A. 2017. Space oddity: the mission for spatial integration. Can. J. Fish. Aquat. Sci. 74(11): 1698-1716.

Beverton, R.J.H., and Holt, S.J. 1957. *On the dynamics of exploited fish populations*. Fisheries Investment Series 2. 19. U.K. Ministry of Agriculture and Fisheries. London: Chapman and Hall. 583 pp.

Brownie, C. Hines, J.E., Nichols, J.D. Pollock, K.H., and Hestbeck, J.B. 1993. Capture-recapture studies for multiple strata including non-Markovian transition probabilities. Biometrics. **49**: 1173-1187.

Cadigan, N. 2016. Updates to a northern cod (*Gadus morhua*) state-space integrated assessment model. DFO Can. Sci. Advis. Res. Doc. 2016/022. v + 58p.

Cadrin, S.X., Goethel, D.R., Morse, M.R., Gay, F., and Kerr, L.A. 2019. "So, where do you come from?" The impact of assumed spatial population structure on estimates of recruitment. Fish. Res. DOI: 10.1016/j.fishres.2018.11.030.

Carruthers, T.R., Walter, J.F., McAllister, M.K., and Bryan, M.D. 2015. Modeling age-dependent movement: an application to red and gag groupers in the Gulf of Mexico. Can. J. Fish. Aquat. Sci. 72: 1159-1176.

Cope, J.M., and Punt, A.E. 2011. Reconciling stock assessment and management scales under conditions of spatially varying catch histories. Fish. Res. 107: 22–38.

de Moor, C.L., Butterworth, D.S., and van der Lingen, C. 2017. The quantitative use of parasite data in multi-stock modeling of South African sardine (*Sardinops sagax*). Can. J. Fish. Aquat. Sci. 74(11): 1895-1903.

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

Goethel, D.R., Quinn, T.J., II, and Cadrin, S.X. 2011. Incorporating spatial structure in stock assessment: movement modelling in marine fish population dynamics. Rev. Fish. Sci. **19**: 119–136.

Goethel, D.R., Legault, C.M., and Cadrin, S.X. 2015a. Demonstration of a spatially explicit, tag-integrated stock assessment model with application to three interconnected stocks of yellowtail flounder off of New England. ICES J. Mar. Sci. **72**: 582–601.

Goethel, D.R., Legault, C.M., and Cadrin, S.X. 2015b. Testing the performance of a spatially explicit tag-integrated stock assessment model of yellowtail flounder (*Limanda ferruginea*) through simulation analysis. Can. J. Fish. Aquat. Sci. **72**(1): 164–177.

Goethel, D.R., and Berger, A.M. 2017. Accounting for spatial complexities in the calculation of biological reference points: effects of misdiagnosing population structure for stock status indicators. Can. J. Fish. Aquat. Sci. 74(11): 1878-1894.

Guan, W., Cao, J., Chen, Y., and Cieri, M. 2013. Impacts of population and fishery spatial structures on fishery stock assessment. Can. J. Fish. Aquat. Sci. **70**:1178-1189.

Hampton, J., and Fournier, D.A. 2001. A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. Mar. Freshwater Res. **52**: 937-963.

Hanselman, D.H., Heifetz, J., Echave, K.B., and 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.

Hoenig, J.M., Barrowman, N.J., Pollock, K.H., Brooks, E.N., Hearn, W.S., and Polacheck, T. 1998. Models for tagging data that allow for incomplete mixing of newly tagged animals. Can. J. Fish. Aquat. Sci. **55**: 1477-1483.

Hulson, P-J.F., Miller, S.E., Ianelli, J.N., and Quinn, T.J., II. 2011. Including tag-recovery data into a spatial age-structured model: walleye pollock (*Theragra chalcogramma*) in the eastern Bering Sea. Can. J. Fish. Aquat. Sci. **68**: 1625–1634.

Hulson, P-J.F., Quinn, T.J., II, Hanselman, D.H., and Ianelli, J.N. 2013. Spatial modelling of Bering Sea walleye Pollock with integrated age-structured assessment models in a changing environment. Can. J. Fish. Aquat. Sci. **70**: 1-15.

ICES (International Council for the Exploration of the Sea). 2017. Report of the benchmark workshop on widely distributed stocks (WKWIDE), 30 January-3 February 2017, Copenhagen, Denmark. ICES CM 2017\ACOM:36. 196pp.

Kerr, L.A., Hintzen, N.T., Cadrin, S.X., Clausen, L.W., Dickey-Collas, M., Goethel, D.R., Hatfield, E.M.C., Kritzer, J.P., and Nash, R.D.M. 2016. Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. ICES J. Mar. Sci. 74(6): 1708-1722.

Kolody, D., and Hoyle, S. 2015. Evaluation of tag mixing assumptions in western Pacific Ocean skipjack tuna stock assessment models. Fish Res. 163: 127-140.

Lauretta, M., and Goethel, D.R. 2017. The robustness of Brownie tag return models to complex spatiotemporal dynamics evaluated through simulation analysis. Can. J. Fish. Aquat. Sci. 74(11): 1845-1861.

Lee, H.-H., Piner, K., Maunder, M., Taylor, I., and Methot, R. 2017. Evaluation of alternative modelling approaches to account for spatial effects due to age-based movement. Can. J. Fish. Aquat. Sci. 74(11): 1832-1844.

Li, Y., Bence, J.R., and Brenden, T.O. 2015. An evaluation of alternative assessment approaches for intermixing fish populations: a case study with Great Lakes lake whitefish. ICES. J. Mar. Sci. **72**(1): 70-81.

Li, Y., Bence, J.R., and Brenden, T.O. 2018. Can spawning origin information of catch or a recruitment penalty improve assessment performance for a spatially structured stock assessment model? Can. J. Fish. Aquat. Sci. 75(12): 2136-2148.

MacCall, A.D. 1990. *Dynamic Geography of Marine Fish Populations*. Washington Sea Grant Program. University of Washington Press, Seattle, Washington.

Maunder, M. 1998. Integration of tagging and population dynamics models in fisheries stock assessment. University of Washington. **PhD:** 306p.

Maunder, M.N. 2001. Integrated tagging and catch-at-age analysis (ITCAAN): model development and simulation testing. In: Kruse, G.H., Bez, N., Booth, A., Dorn, M.W., Hills, S., Lipcius, R.N., Pelletier, D., Roy, C., Smith, S.J., and Witherell, D. (Eds.), Spatial Processes and Management of Fish Populations. Alaska Sea Grant College Program Report, AK-SG-01-02. p. 123-146.

Maunder, M.N., and Punt, A.E. 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. **142**: 61-74.

McGilliard, C.R., Punt, A.E., Methot, R.D., and Hilborn, R. 2015. Accounting for marine reserves using spatial stock assessments. Can. J. Fish. Aquat. Sci. **72**: 262-280.

Punt, A.E. 2019a. Modeling recruitment in a spatial context: a review of current approaches, simulation evaluation of options, and suggestions for best practices. Fish. Res. DOI: 10.1016/j.fishres.2017.08.021.

Punt, A.E. 2019b. Spatial stock assessment methods: a viewpoint on current issues and assumptions. Fish. Res. 213: 132-143.

Punt, A.E., Pribac, F., Walker, T.I., Taylor, B.L., and Prince, J.D. 2000. Stock assessment of school shark, *Galeorhinus galeus*, based on a spatially explicit population dynamics model. Mar. Freshwater Res. 51(3): 205-220.

Punt, A.E., Haddon, M., and Tuck, G.N. 2015. Which assessment configurations perform best in the face of spatial heterogeneity in fishing mortality, growth, and recruitment? A case study based on pink ling in Australia. Fish. Res. **168**: 85-99.

Punt, A.E., Haddon, M., Little, L.R., and Tuck, G.N. 2016. Can a spatially-structured stock assessment address uncertainty due to closed areas? A case study based on pink ling in Australia. Fish. Res. **175**: 10-23.

Punt, A.E., Deng, R.A., Siddeek, M.S.M., Buckworth, R.C., and Vanek, V. 2017a. Data weighting for tagging data in integrated size-structured models. Fish. Res. 192: 94-102.

Punt, A.E., Haddon, M., Little, L.R., and Tuck, G.N. 2017b. The effect of marine closures on a feedback control management strategy used in a spatially aggregated stock assessment: a case study based on pink ling in Australia. Can. J. Fish. Aquat. Sci. 74(11): 1960-1973.

Punt, A.E., Okamoto, D.K., MacCall, A.D., Shelton, A.O., Armitage, D.R., Cleary, J.S., Davies, I.P., Dressel, S.C., Francis, T.B., Levin, P.S., Jones, R.R., Kitka, H., Lee, L.C., McIsaac, J.A., Poe, M.R., Reifenstuhl, S., Silver, J.J., Schmidt, J.O., Thornton, T.F., Voss, R., and Woodruff, J. 2018. When are estimates of spawning stock biomass for small pelagic fishes improved by taking spatial structure into account? Fish. Res. 206: 65-78.

Quinn, II, T.J., R.B. Deriso, and P.R. Neal. 1990. Migratory catch age analysis. Can. J. Fish Aquat. Sci. **47**: 2315-2327.

Rogers, L.A., Storvik, G.O., Knutsen, H., Olsen, E.M., and Stenseth, N.C. 2017. Fine-scale population dynamics in a marine fish species inferred from dynamic state-space models. J. Anim. Ecol. 66: 888-898.

Thorson, J.T., Wetzel, C. 2016. The Status of Canary Rockfish (*Sebastes pinniger*) in the California Current in 2015. Pacific Fishery Management Council, 7700 NE Ambassador Way, Portland, OR 97220. http://www.pcouncil.org/groundfish/stockassessments/by-species/canary-rockfish/.

Vincent, M.T., Brenden, T.O., and Bence, J.R. 2017. Simulation testing the robustness of a multi-region tag-integrated assessment model that exhibits natal homing and estimates natural mortality and reporting rate. Can. J. Fish. Aquat. Sci. 74(11): 1930-1949.

Waterhouse, L., and Hoenig, J.M. 2011. Instantaneous-rates tagging models allowing for delayed mixing of newly tagged cohorts: partial year tabulation of recaptures. N. Am. J. Fish. Manag. 31: 995-1004.

Ying, Y., Chen, Y., Lin, L., and Gao, T. 2011. Risks of ignoring fish population spatial structure in fisheries management. Can. J. Fish. Aquat. Sci. **68**: 2101–2120.

Ziegler, P.E. 2013. Influence of data quality and quantity from a multiyear tagging program on an integrated fish stock assessment. Can. J. Fish. Aquat. Sci. **70**: 1031-1045.

7. Tables

Table 1. Uncertainty associated with each data set used to simulate observation error in the simulation model and input as data weights in the estimation models (including effective sample size (*ESS*) for age composition and tagging data). The variance levels used to simulate recruitment deviations in the operating model and subsequently used to penalize deviations from average recruitment in the estimation model are also provided. *ESS* and variance (σ) are constant across years, while tagging *ESS* is also constant across cohorts (i.e., each tagged cohort has the same *ESS*). Models without tagging data use the same weighting, but have no tagging component.

Data component	Distribution	Parameter	Base settings	
			Population 1	Population 2
Fishery age composition	Multinomial	*ESS*	150.	150.
Fishery yield	Lognormal	σ	0.05	0.05
Survey age composition	Multinomial	*ESS*	150.	150.
Survey biomass index	Lognormal	σ	0.2	0.2
Tagging data	Multinomial	*ESS*	200.	200.
Recruitment variance	Lognormal	σ	0.55	0.5

Table 2. Operating model assumptions and inputs for the tagging sub-model. Descriptions are provided for the *Base* model scenario. Models that differ from the *Base* scenario settings in terms of operating model assumptions are denoted by a (^), whereas those differing in the estimation model assumptions are denoted by a (*). Scenario names are provided in Table 3.

Tag component	Base settings	Models with alternate settings
Total number of tag releases per year	5,000	*Tag_Opp* ^
Population distribution	Proportional to survey abundance	*Tag_Opp* ^, *Tag_Area_2* ^
Age distribution of tags	Proportional to survey age compositions	*No_Age_Tag* *
Frequency of tagging	Every five years	*No_Tag* ^, *Tag_Yrly* ^, *Tag_evy_10* ^, *Tag_Beg_5* ^, *Tag_Mid_5* ^, *Tag_End_5* ^,
Tag lifespan	Five years	
Tag mixing	Fully mixed	*No_Tag_Mx* ^, *Est_Tag_Mx* ^*

Table 3. Description of simulation scenarios, associated scenario abbreviations, critical model assumptions, and convergence rates. OM indicates the operating model and EM represents the estimation model.

Simulation group	Scenario name	Tag design (OM)	Tag assumptions (EM)	Movement parametrization	Other notes	Convergence rate
Tag time series (Group 1)	*Base*	Tag every 5 years	Match OM	Estimate in 2 year time blocks		0.980
	No_Tag	No tagging	Match OM	Estimate in 2 year time blocks	Spatial model with no tagging	1.000
	Tag_Yrly	Tag every year	Match OM	Estimate in 2 year time blocks		0.990
	Tag_Evy_10	Tag every 10 years	Match OM	Estimate in 2 year time blocks		0.990
	Tag_Beg_5	Tag during the first 5 years	Match OM	Estimate in 2 year time blocks		1.000
	Tag_Mid_5	Tag during the middle 5 years	Match OM	Estimate in 2 year time blocks		1.000
	Tag_End_5	Tag during the terminal 5 years	Match OM	Estimate in 2 year time blocks		0.990
Tag deployment (Group 2)	*Opp_Tag*	Tag opportunistically by population and year	Match OM	Estimate in 2 year time blocks	The probability of a tag event occurring is determined by a series of independent Bernouilli and uniform distributions (see Table 3)	1.000
	Tag_Area_2	Tag only in population 2 every 5 years	Match OM	Estimate in 2 year time blocks		1.000
Tag assumptions (Group 3)	*No_Age_Tag*	Tag every 5 years	Do not fit age-based cohorts	Estimate in 2 year time blocks	EM assumes no information on age of tagged fish	1.000
	No_Tag_Mx	Tag every 5 years, assume incomplete mixing of tags	Ignore incomplete mixing	Estimate in 2 year time blocks	Movement and F differ for tagged fish in year of tag release in OM, EM assumes complete mixing	1.000
	Est_Tag_Mx	Tag every 5 years, assume incomplete mixing of tags	Estimate incomplete mixing rates	Estimate in 2 year time blocks	Movement and F differ for tagged fish in year of release in OM, EM estimates these parameters	1.000
Life History (Group 4)	*LL_Tag_Evy_5*	Tag every 5 years	Match OM	Estimate in 2 year time blocks	Long-lived life history (16 ages)	1.000
	SL_Tag_Evy_5	Tag every 5 years	Match OM	Estimate in 2 year time blocks	Short-lived life history (4 ages)	0.890
Movement parametrization (Group 5)	*No_Move*	No tagging	Match OM	No movement estimated	OM includes movement among populations, but EM assumes two closed populations	1.000
	Cnst_Move	Tag every 5 years	Match OM	Estimate constant movement		0.990

Table 4. Mean relative error (MRE) aggregated across all years for important population parameters. Scenario names are from Table 3. System values for biomass and recruitment represent the MRE aggregated across populations. Values for the movement parameters represent the MRE for the estimated movement rates (i.e., emigration not residency). An NA indicates that the value was not estimated in the given scenario.

Simulation group	Scenario name	Biomass			Fishing Mortality		Recruitment			Movement	
		Population 1	Population 2	System	Population 1	Population 2	Population 1	Population 2	System	Population 1 to 2	Population 2 to 1
Tag time series (Group 1)	*Base*	0.61	0.39	0.33	7.05	1.48	1.86	7.16	0.76	6.71	5.49
	No_Tag	-1.36	2.96	0.3	14.63	0.01	1.07	17.07	0.87	71.78	74.86
	Tag_Yrly	0.67	0.11	0.26	7.96	1.88	0.9	7.73	0.73	5.7	6.3
	Tag_Evy_10	0.23	0.52	0.19	9.17	1.37	4.04	6.79	0.65	21.11	23.36
	Tag_Beg_5	0.1	0.85	0.23	9.47	0.65	10.09	-1.22	0.86	42.03	39.08
	Tag_Mid_5	0.35	0.62	0.28	8.69	1.0	8.11	4.64	0.84	40.14	42.45
	Tag_End_5	-0.47	1.75	0.28	10.93	-0.42	7.17	5.74	0.8	53.84	57.23
Tag deployment (Group 2)	*Opp_Tag*	0.07	1.07	0.32	10.62	0.64	5.57	6.51	0.85	25.11	30.47
	Tag_Area_2	-0.87	2.65	0.45	11.38	-0.4	-4.22	18.19	1.04	-3.07	1.67
Tag assumptions (Group 3)	*No_Age_Tag*	3.49	14.3	7.82	-0.91	-14.76	-2.34	31.77	6.32	14.5	16.14
	No_Tag_Mx	8.7	12.8	10.09	-5.99	-14.04	11.12	12.8	6.9	-9.03	-13.17
	Est_Tag_Mx	0.14	1.9	0.74	7.47	-0.21	1.0	10.57	1.15	1.38	5.05
Life History (Group 4)	*SL_Tag_Evy_5*	-0.47	1.1	0.03	2.2	-1.77	5.79	-1.82	0.21	32.51	18.74
	LL_Tag_Evy_5	0.8	1.2	0.76	4.96	3.9	3.81	5.21	0.68	7.54	5.75
Movement parametrization (Group 5)	*No_Move*	-6.54	13.14	1.43	20.57	-12.19	-2.76	32.58	1.73	NA	NA
	Cnst_Move	2.34	-1.85	0.1	5.05	4.16	7.81	-0.11	0.57	16.1	6.67

Table 5. Median absolute relative error (MARE) aggregated across all years for important population parameters. Scenario names are from Table 3. System values for biomass and recruitment represent the MARE aggregated across populations. Values for the movement parameters represent the MARE for the estimated movement rates (i.e., emigration not residency). An NA indicates that the value was not estimated in the given scenario.

Simulation group	Scenario name	Biomass			Fishing Mortality		Recruitment			Movement	
		Population 1	Population 2	System	Population 1	Population 2	Population 1	Population 2	System	Population 1 to 2	Population 2 to 1
Tag time series (Group 1)	*Base*	3.29	4.63	1.47	7.32	6.09	9.37	12.94	3.17	21.93	25.52
	No_Tag	5.23	8.57	1.88	13.67	10.81	14.99	22.96	3.39	72.05	91.04
	Tag_Yrly	2.95	4.17	1.42	8.26	5.58	8.58	11.16	3.16	20.04	20.5
	Tag_Evy_10	3.85	5.75	1.57	8.95	7.53	10.99	15.82	3.24	33.06	38.98
	Tag_Beg_5	4.31	6.62	1.63	9.29	8.56	11.7	17.99	3.31	51.98	55.39
	Tag_Mid_5	4.05	6.03	1.65	8.9	7.76	11.04	15.63	3.29	48.37	57.48
	Tag_End_5	4.0	5.89	1.53	10.17	7.81	12.08	17.48	3.2	54.76	69.15
Tag deployment (Group 2)	*Opp_Tag*	4.22	6.14	1.65	10.48	7.98	11.48	16.65	3.31	44.34	46.62
	Tag_Area_2	4.06	5.94	1.62	11.27	6.78	11.32	16.52	3.28	34.92	26.57
Tag assumptions (Group 3)	*No_Age_Tag*	5.42	13.12	6.35	6.8	15.64	14.47	26.5	4.7	40.99	39.2
	No_Tag_Mx	6.97	12.24	8.38	8.09	15.05	11.13	15.33	4.86	32.4	47.73
	Est_Tag_Mx	3.53	5.14	1.59	7.8	6.58	10.1	14.46	3.25	31.14	31.79
Life History (Group 4)	*SL_Tag_Evy_5*	3.74	5.37	2.11	5.53	7.59	9.34	13.13	3.5	33.17	30.81
	LL_Tag_Evy_5	3.33	4.18	1.26	5.82	5.94	10.6	15.24	3.21	18.83	22.88
Movement parametrization (Group 5)	*No_Move*	6.82	12.11	2.06	19.97	12.87	14.98	23.41	3.4	NA	NA
	Cnst_Move	3.93	5.72	1.48	7.01	7.42	11.74	17.96	3.18	28.39	23.62

8. Figures

Figure 1. Violin plots illustrating the distribution of percent relative error in biomass, recruitment, fishing mortality, and movement by population for all runs within each scenario and across all years in the assessment time series. Overlaid boxplots provide the interquartile range and median (line). Points are the mean values. The vertical dashed line represents zero bias. Scenario names are explained in Table 3.

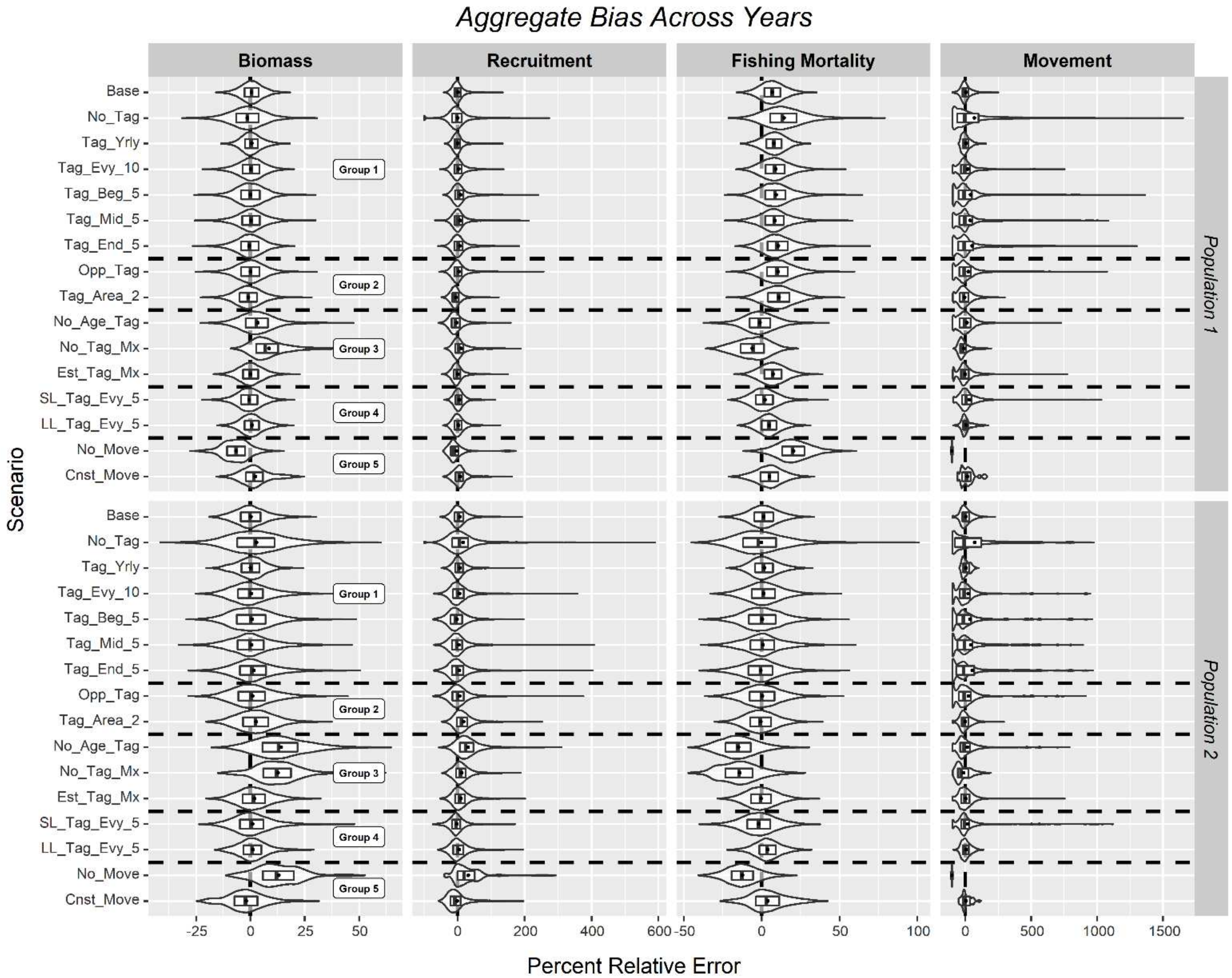

Figure 2. Violin plots illustrating the terminal year (year 30) distribution of percent relative error in biomass, recruitment, fishing mortality, and movement by population for all runs within each scenario. Overlaid boxplots provide the interquartile range and median (line). Points are the mean values. The vertical dashed line represents zero bias. Scenario names are explained in Table 3.

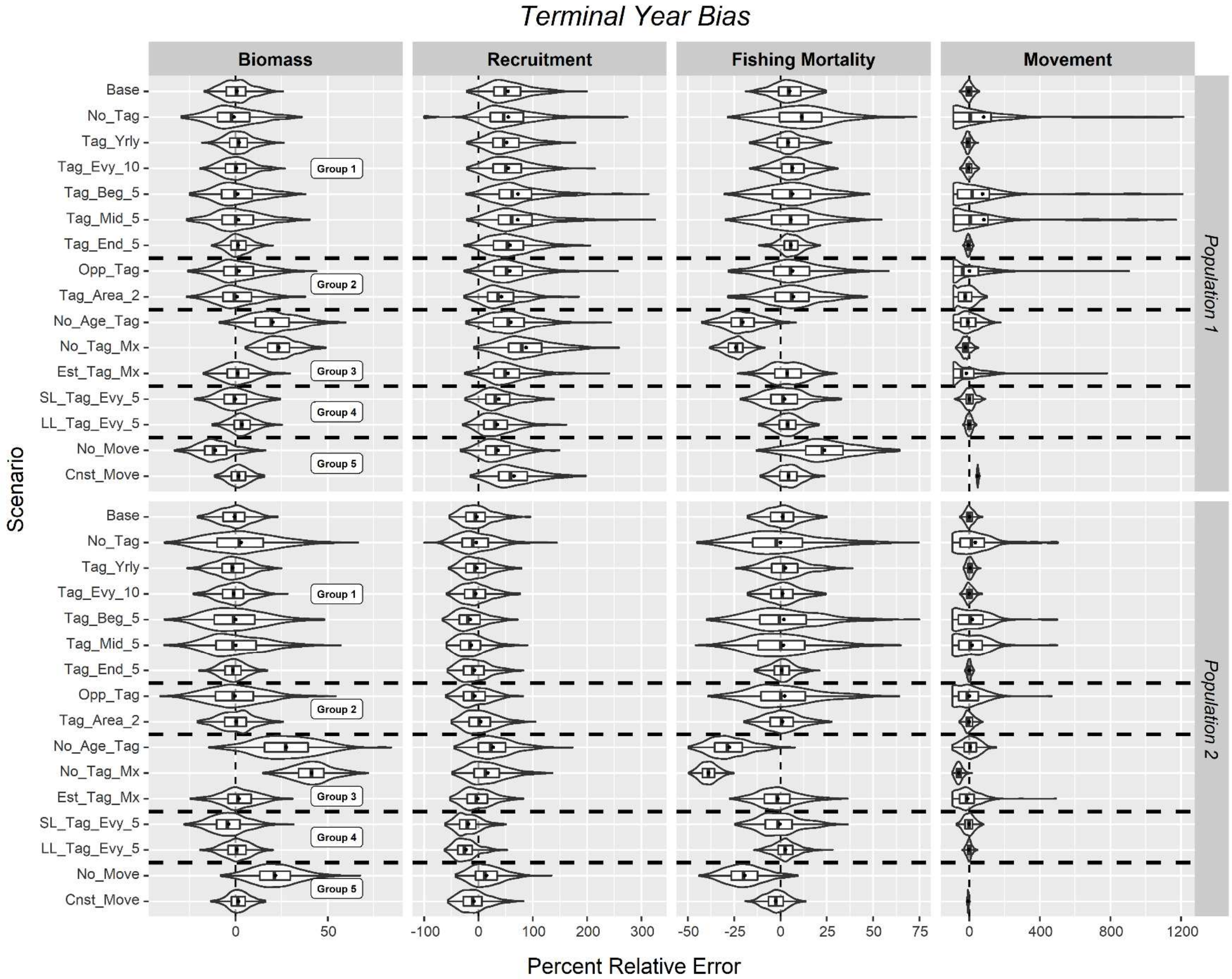

Figure 3. Time series plots illustrating the distribution (violin plots) of movement (left panel), recruitment (middle panel), and fishing mortality (right panel) parameters for population one across several tag time series scenarios. Points represent the annual median estimate across all of the converged models. The line represents the operating model true value. For the movement plot, the black fill and points represent residency in population one (higher proportions in each plot), whereas the grey fill and points represent movement rates to population two (lower proportions in each plot). Scenario names are provided above each plot and are described in Table 3.

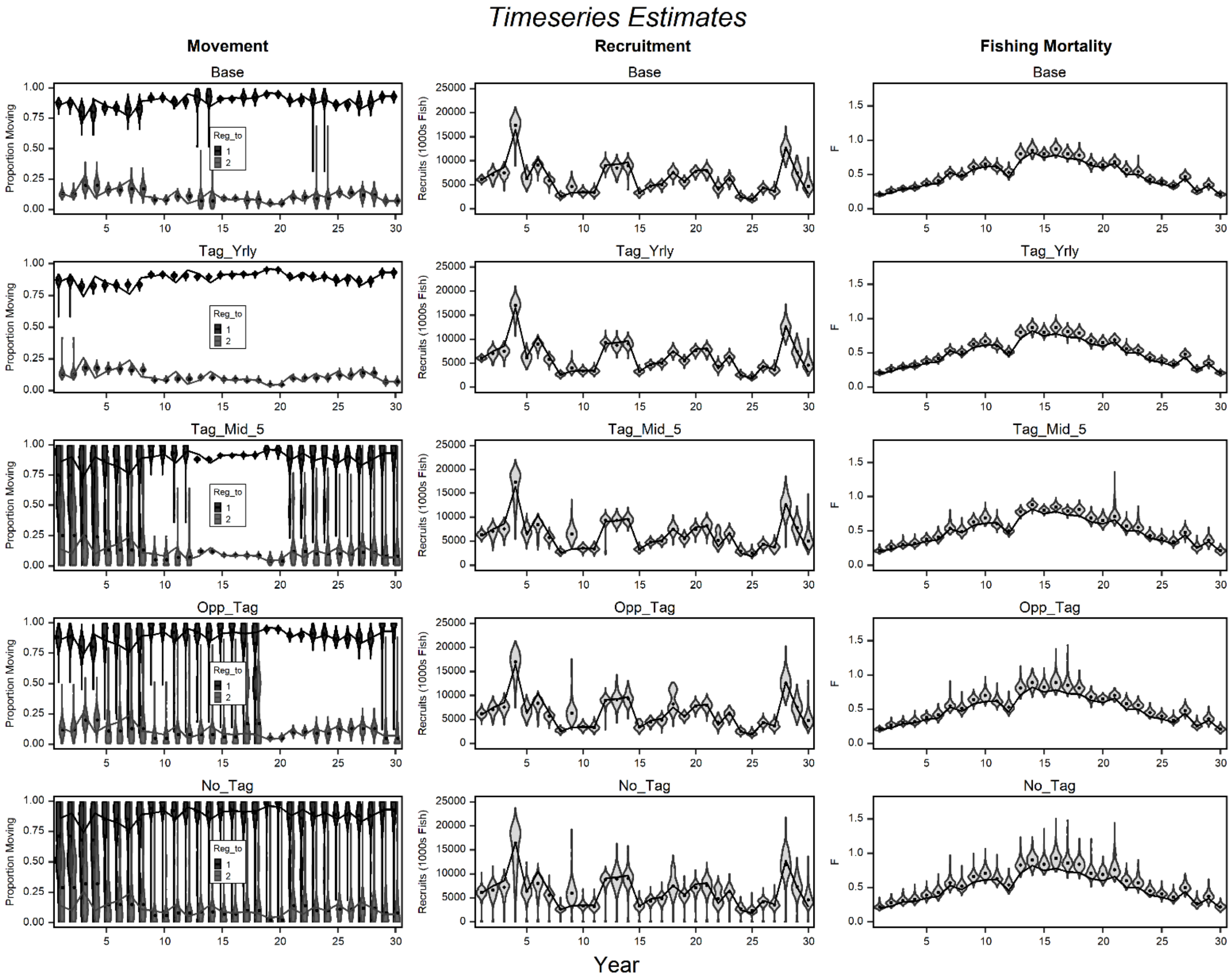

Figure 4. The approximate relative cost of a simulated tagging design plotted against the resulting median absolute relative error (MARE) aggregated across years for important population-specific parameters. Relative cost is estimated based on the frequency of tagging, the number of tags released, and the spatial distribution of tagging. Point labels provide the scenario name as described in Table 3. The results for the *Base* scenario and the scenario that does not include tagging data are highlighted with black label fill. Note that the x-axis scales differ by panel.