

# **Adapting benchmarks of biological status for persistent changes in productivity and variability in exploitation history with a focus on data-limited populations (Conservation Units) of chum salmon in southern BC**

Carrie Holt<sup>1</sup>, Brooke Davis<sup>1</sup>, Lyse Godbout<sup>1</sup>, Pieter Van Will<sup>2</sup>, and Wilf Luedke<sup>3</sup>

<sup>1</sup>Pacific Biological Station, DFO, Nanaimo, BC; <sup>2</sup>DFO, Port Hardy, BC; <sup>3</sup>DFO, South Coast Area Office, Nanaimo, BC

## **Abstract**

Canada's Wild Salmon Policy requires the biological assessment of conservation units (CUs) of Pacific salmon to ensure their conservation for future generations. A "stop light" approach has been adopted that uses stock-recruit information and benchmarks to assign green, amber, or red status to CUs. Data limitations for many CUs require the exploration of revised benchmarks to ensure conservation objectives are achieved when stock-recruitment data are not available. In this study we compare the performance of revised benchmarks that consider variation in productivity and exploitation rates using prospective simulation modelling and retrospective analyses of empirical data for chum salmon in southern BC. In retrospective analyses, we found that benchmarks based on percentiles of escapement time series are generally more precautionary than previously adopted stock-recruit based benchmarks for the seven CUs of chum salmon analyzed here. The simulation study yielded similar results. In most cases percentile benchmarks were a precautionary choice to reach conservation objectives. However, when population productivity was low and harvest rates high, neither benchmark type was precautionary and percentile-based benchmarks were especially risky. Data truncation methods that aim to adapt benchmarks to current productivity regimes and exploitation rates were variably effective, depending on the benchmark type and current population dynamics, and should therefore be considered carefully on a case-by-case basis.

## **Introduction**

The Pacific Salmon Treaty (PST) Chum Annex requires biological benchmarks to inform the development of fishery reference points for PST related fisheries; including the lower fishery reference point for the Johnstone Strait fisheries and subsequent terminal fisheries. Biological benchmarks for data-limited populations have been proposed and are currently being applied to Conservation Units (CUs; population units of biological assessment under Canada's Wild Salmon Policy) of chum salmon in southern BC. However, recent simulation modelling has shown that these benchmarks, which are derived from time-series of spawner abundance alone, are associated with high probabilities of extirpation under low and/or declining stock productivity (Holt and Folkes 2015). In addition, previous unpublished studies suggest these benchmarks may be higher than necessary to achieve conservation objectives when exploitation rates are low (cited in Fair et al. 2010).

Reference points that are currently being used for management are 20-35 years out of date, and do not reflect current trends in productivity, stock status, or other ecosystem considerations. Benchmarks of biological status (and revised versions developed here for southern BC CUs) can be used to inform reference points and resulting management decisions (Holt and Irvine 2013).

Certain fisheries in both countries are known to impact chum salmon originating from the other country (Pacific Salmon Commission Joint Chum Technical Committee 2013).

To address these gaps, our original research objectives were to (1) identify candidate revised benchmark(s) for data-limited CUs (having time-series of relative abundance of spawners only) that account for persistent changes in productivity and/or variability in exploitation rate history, and (2) compare performance of revised benchmarks that account for changes in productivity and variability in exploitation rate history for data-limited CUs against benchmarks derived from data-intensive methods using prospective, simulation modelling, and retrospective analyses of empirical data.

However, the methods proposed to address these objectives include comparisons of benchmarks derived for data-limited with data-rich scenarios, and require assumptions about data quality. Given pervasive observation errors, inconsistent data, and uncertainties in assessments, estimates of benchmarks will differ from underlying “true” values. Previous evaluations of the relative performance of benchmarks have not fully accounted for these data uncertainties. Therefore, we revised the first objective to evaluate benchmarks given uncertainties in underlying data. In addition, we focused our evaluation of methods for accounting for productivity changes to a data truncation approach (objective 2). This approach uses subsets of time-series data to estimate benchmarks when productivity or capacity has changed over time.

Our revised objectives were:

- (1) Identify relative performance of benchmarks for data-limited CUs that use either spawner-recruitment data or time-series of spawner abundances alone (percentile-based benchmarks) by assessing how well they track benchmarks derived from the “true” underlying stock-recruitment parameters assuming perfect knowledge. Performance is evaluated in (a) retrospective analyses of empirical data and (b) a simulation model of a hypothetical CU.
- (2) Use simulation analyses to compare percentile-based benchmarks that have been revised to account for changes in capacity and/or productivity against benchmarks derived from data-intensive methods, where revised benchmarks are calculated by limiting time-series to either historical high-production (productivity) or current low-production (productivity) regime.

## **Canada’s Wild Salmon Policy and biological benchmarks**

Canada’s Wild Salmon Policy (2005) outlines strategies to ensure the conservation of wild Pacific Salmon for future generations. The policy requires the biological assessment of CUs into one of three status zones: green, amber and red. The lower benchmark, delineating red and amber zones, is to be established at a level ensuring the CUs is buffered from being considered at risk of extinction under COSEWIC, the Committee on the Status of Endangered Wildlife in Canada, taking into account data uncertainties and harvest management. The upper benchmark, delineating amber and green zones is the escapement level associated with the maximum average annual catch, under current environmental conditions. While this policy lays out a basic framework for the assessment of conservation status of CUs, it does not require a single set of

benchmarks for all CUs. Rather, it states that benchmarks will be determined on a “case-by case basis, and depend on available information and the risk tolerance applied” (DFO2005).

For populations with time-series of stock-recruit data, benchmarks were identified by Holt et al. (2009) to be robust to uncertainties in underlying stock productivity. These benchmarks are based on the Ricker stock-recruit relationship, which is widely used for Pacific salmon populations (Ricker 1975). The lower benchmark,  $S_{gen}$ , is the number of spawners required to rebound to  $S_{MSY}$  within one generation, under equilibrium conditions, in the absence of fishing. The upper benchmark is 80% of  $S_{MSY}$ , the number of spawners required to achieve maximum sustainable yield (MSY). Alternatively, for those CUs with limited or uncertain stock-recruit data, alternative benchmarks are being developed. Percentile-based approaches have been proposed for determining sustainable escapement goals (SEGs) or conservation benchmarks under Canada’s Wild Salmon Policy (Clark et al. 2014, Holt and Folkes 2015). These methods require escapement data only, and simply compare current escapement levels with the percentiles of historical observations. The Alaska Department of Fish and Game (ADF&G) compared various percentiles as a basis for SEGs (intended to approximate  $S_{MSY}$ ) in a simulation evaluation and retrospective analysis (Clark et al. 2014). Based on this work, a multi-tier system was recommended, where percentile values for SEGs are chosen based on data contrast, data uncertainty, and harvest rates. In particular, Clark et al. (2015) recommend that percentile-based SEGs not be used when harvest rates are high ( $> 40\%$ ), or spawner and recruitment data show little contrast over time, and measurement error is high. Our evaluation differs from Clark et al. (2015) in that we evaluated the extent to which percentile-benchmarks are consistent with biological benchmarks already identified under the Wild Salmon Policy, instead of applying (and evaluating) them as escapement goals for management at MSY levels. For data-limited chum salmon CUs in southern BC, percentile benchmarks at the 25<sup>th</sup> and 75<sup>th</sup> percentiles have been proposed and provisionally implemented as lower and upper benchmarks, respectively (Hilborn et al. 2013). These percentiles are higher, and therefore more conservative, or precautionary, than SEGs proposed by the ADF&G.

In this report, we first provide Methods and Results for Objective 1(a), evaluating data-limited benchmarks using retrospective analyses, followed by Objective 1(b), evaluating data-limited benchmarks in simulation (Objective 1b). We then describe Methods and Result for Objective 2, evaluating benchmarks that use truncated time-series data to account for changes in productivity, and provide a synthesis Discussion for both objectives.

## **Objective 1a. Evaluating benchmarks using retrospective analyses**

In retrospective analyses, we compared biological status of 7 CUs of Inner South Coast chum salmon using data-rich and data-limited benchmarks derived from historical time-series data. Escapement data is available for these CUs from 1953-2012, while CU-specific return data was reconstructed from exploitation rates, migration timing and patterns, spawner abundances, and age distributions, for brood years 1955-2006 (P. van Will pers. comm. 2016). Run reconstructions use catch data and information about migration timing and patterns of fish from specific CUs through different fisheries to estimate the number of returning fish originating from each CU. However, CU-specific return estimates are very sensitive to uncertainty in migration timing and patterns (see Korman J. et al. 2013). In years where spawner abundances were missing, data have been infilled using standard approaches (Van Will 2014). On average, across CUs and years, 45% of sampling sites were surveyed (ranging from 27% for Howe Sound – Burrard Inlet to 57% for Bute Inlet). Infilling occurred at the CU level for 2 CUs in years where no sites were surveyed (17 of 61 years for the Upper Knight CU and 8 of 61 years for the Bute Inlet CU). Fitting the Ricker model to uncertain data can lead to biased parameter estimates because of observation errors-in-variables and time-series biases (Walters and Martell 2004). Time-series are relatively long (51 years) and contrast in escapement observations is relatively high (ratio of maximum to minimum spawner abundances ranged from 8-2600, mean=481), which may ameliorate these biases. However, caution in the interpretation of results is warranted. These results should be considered with those from simulation model that incorporates multiple sources of data uncertainties (Objective 1b).

For the data-rich scenarios, we compared benchmarks derived from two different forms of the Ricker model, the standard model which estimates parameters independently for each CU, and a hierarchical model where CUs are assumed to have similar productivities. Hierarchical models may reduce uncertainties and biases in parameter estimation mentioned above by sharing information on productivity across CUs. CU-specific productivities were drawn from a global hyper-distribution given evidence for spatial covariation in productivity among populations within regions (Pyper et al. 2002). Following the results of Holt and Folkes (2015) who investigated the impacts of temporal changes in productivity, we further identified temporal trends in productivity over time in these CUs using a recursive Bayes modelling approach.

## **Methods**

Stream-specific escapement for inner south coast chum were aggregated to the CU level, and identified as either wild, or enhanced (hatchery-origin fish, or those fish used for hatchery brood stock). Wild escapement were infilled at the stream level and then again at the CU level when there were no escapement estimates for a site within a given CU or a CU within the inner south coast region. Infilling assumed that sites within CUs, and CUs within the region contributed their geometric average proportion of overall escapement in years when data were missing. Infilling occurred at the CU level for two out of seven CUs: Upper Knight (1979, 1980, 1982, 1984, 1989, 1991, 1996, 2004-2013) and Bute Inlet (2005, 2006, 2008-2013). CU-specific returns were estimated for all fish using backwards catch reconstructions with variable vulnerability levels for each CU to each fishery (Van Will 2014). To estimate wild returns, we applied the same proportion of wild fish in escapement to catches, i.e., we assumed that enhanced and wild fish

were equally vulnerable to the fishery. Brood year returns were calculated assuming annual estimates of age-at-maturity from the mixed-stock fishery in Johnstone Strait (Van Will 2014).

We first identified benchmarks and assessed status in the most recent year using all available data. The retrospective analysis was then carried out by sequentially calculating benchmarks using all available data up to a given year. For the percentile benchmarks, we assumed that 10 years of data were required to estimate the first benchmark, and benchmarks were re-estimated every year after that (1963-2012). Since recruitment information is required for the stock-recruitment based benchmarks, and recruitment from a given brood year cannot be calculated until the oldest age class has recruited to the fishery, data used to calculate Ricker benchmarks lag behind percentile benchmarks by 6 years. Therefore, stock-recruitment based benchmarks and status are calculated for years 1970-2012, with Ricker models using data from brood years 1964-2006. Benchmarks were compared to generational mean escapement to determine status. Generational mean escapement was estimated as the four-year running geometric average.

### Standard Ricker Model

For each year with sufficient data, a Ricker model (Eqn. 1) was fit in a Bayesian context, using Markov Chain Monte Carlo (MCMC) methods.

$$(1) R = \alpha S e^{-\beta S},$$

where  $R$  is the abundance of adult recruits from a given spawning event,  $S$  is the number of spawners that generated those recruits (also referred to as escapement). The parameter  $\alpha$ , or productivity, is recruits per spawner at low spawner abundances, and  $\beta$  is the reciprocal of the number of spawners that produce maximum recruits ( $S_{Max}$ ). We linearized the equation and incorporated normally distributed process error, where  $\tau_v$  represents precision of process error and precision is the reciprocal of variance,

$$(2) R = \log(\alpha) + \log(S) - \beta S + v, \quad v \sim \text{normal}(0, \tau_v).$$

We put a weakly informative prior on  $\alpha$  to ensure values greater than zero and within the bounds of observed productivity values for chum salmon (Dorner et al. 2008) (Fig. 1),

$$(3) \log(\alpha) \sim \text{normal}(1, 1).$$

The prior for beta was set indirectly by applying a prior on its reciprocal,  $S_{Max}$ . We had no prior information on  $S_{Max}$ , so we applied a uniform distribution bounded by 1 and twice the maximum observed value (Eqn 4a). In a sensitivity analyses, we also considered a diffuse log-normal distribution for the prior (Eqn 4b), where  $\tau_s$  is the precision of the log-normal prior, calculated using a standard transformation of the coefficient of variation, CV, in normal space to log-normal space. See Appendix for details on the parameterization of priors on  $S_{max}$ .

$$(4a) S_{max} \sim \text{uniform}(1, \max(S_{obs}) * 2)$$

$$(4b) S_{max} \sim \text{lognormal}(\log(\text{mean}(S_{obs})), \tau_s), \quad \tau_s = 1/\log(CV^2 + 1)$$

Uninformative gamma priors were used for  $\tau$  parameters:

$$(5) \tau_v, \tau_S, \sim \text{gamma}(0.01, 0.001)$$

### **Hierarchical Ricker Model**

We estimated Ricker parameters using a hierarchical version of the standard Ricker model (Eqns 1 and 2), where parameters from each CU,  $i$ , were estimated simultaneously. CU-specific  $\alpha_i$  values were drawn from a common, normal distribution:

$$(6a) R = \alpha_i S e^{-\beta_i S} e^v, v \sim \text{normal}(0, \tau_v),$$

$$(6b) \alpha_i \sim \text{normal}(\mu_\alpha, \tau_\alpha),$$

where  $\mu_\alpha$  is the mean of the normal distribution and  $\tau_\alpha$  is precision.

The same prior distributions were used as for the standard Ricker model (Eqns 3-5), with the addition of a prior on the global mean and variance of alpha,  $\mu_\alpha$ .

$$(7) \log(\mu_\alpha) \sim \text{normal}(1, 1)$$

To impose an uninformative prior on  $\tau_\alpha$  we put an uninformative prior on variance  $\sigma_\alpha$ , where  $\sigma_\alpha = 1/\tau_\alpha$ :

$$(8) \sigma_\alpha \sim \text{Uniform}(0, 100)$$

Models were fit using MCMC run using JAGS (Plummer 2003), interfaced through R version 3.2.0 (R Development Core Team 2016) using package “R2jags” (Su and Yajima 2012). Model convergence was assessed using Gelman-Rubin statistics and visual inspection of trace plots.

### **Benchmarks**

For Ricker-based benchmarks, the lower benchmark,  $S_{gen}$ , was calculated numerically, according to the following equation (Holt et al. 2009):

$$(9) S_{MSY} = S_{gen} \alpha e^{-\beta S_{gen}}$$

The upper benchmark was calculated using an approximation developed by Hilborn and Walters (1992):

$$(10) 0.8 S_{MSY} = 0.8 \frac{\log(\alpha)}{\beta} (0.5 - 0.07 \log(\alpha))$$

Percentile benchmarks were calculated as the 25<sup>th</sup> and 75<sup>th</sup> percentile of observed spawner abundances ranked from lowest to highest, for the lower and upper benchmarks respectively ( $S_{25th}, S_{75th}$ ).

### **Changes in productivity**

To identify changes in productivity over time for chum salmon CUs on the inner south coast and assess how those changes affect benchmark performance, we fit a recursive Bayes model to

stock-recruitment data, which allowed for  $\alpha$  to vary over time for each CU individually (Malick and Cox 2016). We fit this model using all available data for each site. It follows the standard Ricker form with a time-varying  $\alpha$  parameter,

$$(11) R = \alpha_t S e^{-\beta S} e^v, v \sim \text{normal}(0, \tau_v),$$

where  $\alpha_t$  is productivity in brood year  $t$ . The model assumes that  $\alpha$  changes over time following a Gaussian random walk:

$$(12) \log(\alpha_t) = \log(\alpha_{t-1}) + w, w \sim \text{normal}(0, \tau_w)$$

The same prior distributions were applied as for the standard Ricker model (Eqns. 3-5), with the addition of a normally distributed prior on  $\alpha$  in year 1, and a uniform prior on the variance associated with the Gaussian random walk  $\sigma_w$ , where  $\sigma_w = 1/\tau_w$ :

$$(13a) \log(\alpha_1) \sim \text{normal}(1, 1), \text{ and}$$

$$(13b) \sigma_w \sim \text{Uniform}(0, 100).$$

## Results

### *Effect of priors on parameter estimates*

Using the standard Ricker model, estimates of  $S_{max}$  were slightly lower when a weakly informative lognormal prior was used for  $S_{max}$  compared with uniform prior, but these differences were small and estimates consistently fell within the range of uncertainty under the alternate assumption (Fig. 2 shows the most recent parameter estimates, 2012). We found that uncertainty in estimates of  $\alpha$  and  $S_{max}$  were reduced slightly in most cases in retrospective analysis of the hierarchical model compared with the standard Ricker model, as expected. The hierarchical Ricker model stabilized parameter estimates for those sites and years with high uncertainty, resulting in slightly narrower credible intervals (Fig. 2).

### *Current benchmarks and status*

Lower percentile benchmarks ( $S_{25th}$ ) tended to be similar in value to lower Ricker-based benchmarks ( $S_{gen}$ ), whereas upper percentile benchmarks ( $S_{75th}$ ) were generally much higher than the Ricker-based upper benchmarks (80%  $S_{MSY}$ , Fig. 3, Table 1). Stock-recruitment benchmarks varied slightly between the standard and hierarchical Ricker models (comparing Fig. 3 (i) vs .(ii) for each CU (a)-(g)), but these differences were small compared with large uncertainties in benchmark estimates (Table 2). The posterior densities of the upper and lower benchmarks overlapped and, in some cases, were nearly indistinguishable, e.g., Southern Coastal Streams and North East Vancouver Island (Fig. 3 (a) and (b)).

As expected, uncertainty in benchmarks was slightly reduced for the hierarchical Ricker model compared with standard Ricker for most CUs (e.g. Southern Coastal Streams, Fig. 3(a)). In cases where uncertainty in benchmarks was higher for the hierarchical model (e.g., Upper Knight, Fig. 3(c), productivity information from neighbouring CUs differed from CU-specific signals, resulting in larger uncertainty.

Statuses for the most recent year, 2012, determined using all data available up to return year 2012 (brood years 1955-2006) are shown in Table 3. Percentile-based status was the same or more precautionary than Ricker-based status in this year. For the two Ricker-based benchmark models, final status matched for all CUs.

### ***Retrospective analyses***

In retrospective analyses, percentile benchmarks tended to vary more over time than Ricker-based benchmarks (Fig. 4). Stock-recruitment benchmarks tended to remain relatively consistent over time for two CUs (Upper Knight, Loughborough), exhibited divergent trends between upper and lower benchmarks (Southern Coastal Streams), or increased over time (North East Vancouver Island, Bute Inlet, Georgia Strait, and Howe Sound to Burrard Inlet). The standard Ricker and hierarchical Ricker benchmarks were nearly indistinguishable from each other over time (comparing first and second row of panels Fig.4(i) and 4(ii)). Uncertainties in stock-recruitment benchmarks tended to decline over time for 3 CUs (Southern Coastal Streams, Upper Knight, and Loughborough), remained approximately consistent over time for three CUs (North East Vancouver Island, Georgia Strait, and Howe Sound to Burrard Inlet), and increased and then declined for Bute Inlet.

For three CUs, percentile benchmarks tended to decline over time (i.e., become less precautionary, Southern Coastal Streams, North East Vancouver Island, and Upper Knight); the others remained constant (Loughborough) or increased over time (Bute Inlet, Georgia Strait, and Howe Sound to Burrard Inlet). The observed declines in percentile benchmarks for 3 CUs were associated with declines in abundance over the entire time series (Southern Coastal Streams), or just the beginning of the time series (North East Vancouver Island and Upper Knight). Although percentile benchmarks decreased over time for some CUs, they tended to be higher (i.e., more precautionary) than stock-recruitment benchmarks.

Statuses varied over time and among methods used to derive benchmarks. Uncertainty in stock-recruitment benchmarks resulted in differences in status when upper or lower credible intervals (calculated using 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of posterior distributions) on benchmarks were used to derive status instead of best estimates (Fig 5, light bars above and below colored status bars for Ricker benchmarks). For example, for Northeast Vancouver Island, in the early 2000's, the assessed status was amber based on best estimate of the standard Ricker benchmarks, but green based on the upper credible interval and red based on the lower credible interval of those benchmarks (Fig.5a).

### **Comparing Benchmarks**

Percentile benchmarks were found to provide the same, or more precautionary status compared to Ricker-based benchmarks (Fig. 5, Tables 3,4). The proportion of years where the two types of benchmarks gave the same status varied across CUs, but averaged 37 and 39% for the standard Ricker and hierarchical Ricker model, respectively (Table 4). On average, the percentile benchmark provided the same or more precautionary status in 93% of years for both model types (Table 4). The relatively few years when percentile benchmarks were lower (less precautionary) than stock-recruitment based benchmarks were associated with either periods of consistently low escapement, declining  $S_{25th}$  benchmarks, and relatively constant  $S_{gen}$  values (e.g., Upper Knight from 1999-2001, see Fig. 3c), or with an abrupt increase in escapement, productivity, and  $S_{gen}$



values, and relatively consistent or slowly increasing  $S_{25th}$  values (e.g., Bute Inlet 1991, 1999-2000, Fig. 3e).

The two stock-recruitment based benchmarks gave the same status 97% of years when averaging across-CUs.

## **Productivity over Time**

Time-varying productivity, estimated using a recursive Bayes model for each CU, showed unique patterns among CUs (Fig. 6). Declines in productivity over time were observed in two CUs (Southern Coastal Streams and Loughborough, Fig. 6a and d), increases followed by declines in three CUs (North East Vancouver Island, Bute Inlet, and Georgia Strait, Fig. 6b, e, and f) and consistent levels followed by a small increase in Howe Sound to Burrard Inlet (Fig. 6g). Estimates of productivity for Upper Knight (Fig. 6c), were highly variable and uncertain. There was considerable uncertainty in estimating  $\alpha$  over time for all CUs, indicated by wide error bounds. Using data to estimate stock-recruitment parameters and benchmarks that spans decades where  $\alpha$  has changed considerably may lead to poor Ricker model fits and large uncertainty in parameter estimates.

## **Objective 1b. Evaluating benchmarks using simulation analyses**

### **Methods**

We adapted the simulation model of Holt and Folkes (2015) to evaluate data-limited benchmarks. As in Holt and Folkes (2015), the model included five components representing population dynamics, observations of abundances, management (including the derivation of benchmarks), harvest, and performance evaluation (Fig. 7). In particular, the model included natural variability in adult recruitment based on a Ricker spawner-recruitment relationship with variable age at maturity, errors in observations of abundances, assessments of biological status relative to benchmarks, the application of a harvest control rule, and uncertainties in the outcomes from implementing management decisions. See Holt and Folkes (2015) for model equations.

Our model differed from that of Holt and Folkes (2015) in 8 ways:

- The population dynamics sub-model included covariance in Ricker residuals among subpopulations within a CU, instead of assuming sub-populations varied independently.
- The observation sub-model was more realistic in that catches (or, alternately exploitation rates) were observed with observation errors, and recruitment by brood year was then calculated using estimated ages-at-maturity, instead of applying observation error directly to “true” recruits by brood year. Annual observation errors in age-at-maturity were simulated using a multivariate logistic distribution (as in natural variability in age-at-maturity).
- In the observation sub-model, we evaluated scenarios where spawner abundances were observed with a consistent negative (or positive) bias that was not corrected for in the assessment.
- In the assessment sub-model, we evaluated scenarios where only a portion of subpopulations were sampled within a CU and a constant expansion factor was applied to

derive escapement estimates for the entire CU. The expansion factor was estimated from observed complete sampling in a 3-year initialization period.

- The assessment sub-model focused on percentile-based benchmarks and stock-recruitment benchmarks applied under the Wild Salmon Policy ( $S_{\text{gen}}$  and 80% of  $S_{\text{MSY}}$  for the lower and upper benchmarks, respectively)
- In the harvest sub-model, a constant low harvest rate (20%) was applied instead of a harvest control rule with limit and/or target reference points. The harvest rates applied over the initialization period were varied in a sensitivity analysis (but remained constant over that period) to reflect the different harvest rate histories observed among CUs.
- In the performance module, benchmarks were evaluated based on the deviations between benchmark estimates and the “true” underlying values. In the case of percentile benchmarks, we evaluated deviations between 25<sup>th</sup> percentile (lower benchmark) and the “true”  $S_{\text{gen}}$  value, and between the 75<sup>th</sup> percentile (upper benchmark) and the “true” 80% of  $S_{\text{MSY}}$  value. Specifically, we evaluated, mean percent error and mean raw error because we were interested in the direction of bias (i.e., if the estimated benchmark was above or below the “true” benchmark) which are reflected in these metrics. We focused our results on mean percent error, MPE, as this metric is scale independent, making comparisons in sensitivity across benchmarks more intuitive. Results for mean raw error are provided in the Appendix.
- The model was run over 50 years, instead of 100 to provide a more realistic time-series length for estimating benchmarks. The model was run over 5000 MC trials, the number of trials required to stabilize output metrics at (standard error  $\leq 3\%$  in performance metrics). The model was initialized for 20 years after a 5-year pre-initialization period necessary to generate the first recruitment by brood year.

## Parameterization

The population dynamics sub-model was parameterized based on previous empirical studies in the primary literature and governmental reports on chum salmon, or other species of Pacific salmon where data on chum were not available. See Holt and Folkes (2015; Appendix) for model equations. The productivity parameter of the spawner-recruitment relationship,  $a$  (defined as  $\log_e(\text{recruits/spawner})$  at low spawner abundance, and referred to simply as productivity here) and the range considered in sensitivity analyses (Table 5, see details below) were chosen to bound productivities observed for six chum salmon stocks from across BC (Dorner et al. 2008; ranging from 0.99-1.94), and three stocks in the Skeena watershed, BC (Korman J. et al. 2013; ranging from 0.7-1.05). Productivity and spawner abundances at equilibrium abundances,  $S_{\text{eq}}$  (set at 10 000 fish) were assumed to be equal among subpopulations.

We assumed an autocorrelation coefficient of 0.6, based on coefficients estimated for three CUs of chum salmon (ranging 0.54-0.68) from Skeena River, BC (Korman et al. 2013), and considered a range of plausible autocorrelation coefficients (0 and 0.9) in sensitivity analyses (Table 5). The standard deviation in recruitment residuals (in log-space) was set to 0.75, within the range of values estimated from the same Skeena River, BC data (0.68-0.90), and within the range estimated for sockeye salmon in BC and Alaska (Korman et al. 1995, Peterman et al. 2003). The average proportions of mature adults at ages 3, 4, and 5 were estimated for 22 chum salmon stocks in BC and Alaska (0.24, 0.67, and 0.09, respectively, Pyper et al. 2002). The

variance in the proportions of ages at maturity was estimated from empirical time series data for age-specific returns of chum salmon in southern BC (1959-2012; Johnstone Strait test fishery and commercial harvest to Statistical Area 12; P. Van Will pers. comm. 2016). The probability of straying among adult recruits was set at 5% based on a review of published stray rates for chum salmon in British Columbia (McElhany et al. 2000).

In the observation sub-model, we assumed the standard deviation in estimates of spawner abundances around the true values (observation errors) was equal to 0.5 (in log-space), which corresponds to an upper estimate of the uncertainty in spawner abundance derived from various visual surveys of Pacific salmon (Cousens 1982, Szerlong and Rundio 2008). Chum salmon abundance is largely estimated from visual surveys, which typically produce relatively imprecise estimates of abundances. We also considered a lower estimate of observation errors of 0.2 in a sensitivity analysis. In the absence of quantitative estimates of uncertainty in catch estimates (commercial, recreational and First Nations subsistence catch), we assumed the same standard deviation in observed catch (0.5 in log-space), and a sensitivity analysis with a lower estimate of 0.2. Although errors in observations of commercial catch are likely less than observation errors in spawner abundance, uncertainties in reporting and estimation of recreational and subsistence harvest are relatively high (Collie et al. 2012, Fleischman et al. 2013).

The standard deviation of outcome uncertainty was estimated at 0.3 using methods described in Collie et al. (2012) by modelling the relationship between catch and total recruitment from two DFO Fishery Statistical Areas of chum salmon on the west coast of Vancouver Island, BC (Dobson et al. 2009). Because the standard deviation of outcome uncertainty is not widely estimated for Pacific salmon and likely varies widely among stocks and management approaches, we also considered an upper value of 0.5 in a sensitivity analysis.

### **Sensitivity analyses**

To assess the strength and direction of effects of input parameters on benchmark performance (measured as deviations between estimated lower benchmarks and “true” lower benchmarks), we performed a sensitivity analysis where each input parameters were varied individually while all others were held constant (Table 5). These sensitivity analyses did not assess the sensitivity of performance to interactions among input variables. For two parameters that had a relatively large effect on performance, productivity and initial harvest rates, a bivariate sensitivity analysis was performed to assess their combined effect on benchmark performance. To further consider interactions among all input variables, we performed a global sensitivity analysis, using the Morris method. Similar to univariate analyses, the Morris method varies each input parameter one at time, but in contrast to univariate analyses, this is done at different points of the factor input space (i.e., at different combinations of other variables) (Morris et al. 2014). The mean elemental effect of an input parameter from the Morris method is an index of the sensitivity of benchmark performance to uncertainty in that parameter. The standard deviation of the elemental effects is an index of sensitivity of benchmark performance to interactions of that variable with other variables. The Morris method was run using the R package, sensitivity, v.1.11.1 (Pujol et al. 2015). We focused univariate and global sensitivity analyses on lower benchmarks (25<sup>th</sup> percentile and  $S_{\text{gen}}$ ), but also considered sensitivity of upper benchmarks (75<sup>th</sup> percentile and 80%  $S_{\text{MSY}}$ ) in our bivariate sensitivity analysis.

### **Results**

Simulation model outputs for an example CU, Southern Coastal Streams are presented in Fig. 8. Harvest rates during the initialization period were drawn at random from the historical time-series of exploitation rates for the Southern Coastal Streams CU, and the productivity parameter was estimated from the historical data for that CU using a hierarchical Ricker model (from Objective 1a). Mean percent error between estimated and “true” benchmarks was generally greater than zero, especially for percentile benchmarks (Fig 8, right panel). Percentile benchmarks tended to be more precautionary than stock-recruitment based benchmarks (i.e., positive deviations were greater), but both were precautionary. Estimates of stock-recruitment benchmarks differed from the “true” values because estimates were based on observed data (black line in Fig 8, left panel, and solid dots in Fig. 8, middle panel) rather than “true” data (grey line in Fig 8, left panel, and hollow dots in Fig. 8, middle panel). The assessed stock-recruitment model (black curve, Fig. 8, middle panel) differed from the “true” underlying model (grey curve, Fig. 8, middle panel) due to those errors in spawner abundance and time-series biases (Walters and Martell 2004).

Performance of lower benchmarks (both  $S_{\text{gen}}$  and  $S_{25\text{th}}$ ) was more sensitive to uncertainty in productivity than to other input parameters (Fig. 9 a and b, respectively). Low productivity values (leftmost black bar) were associated with negative deviations from the base case (i.e., benchmarks that were less precautionary than the base case); high productivities (leftmost white bar) were associated with positive deviations (i.e., benchmarks that were more precautionary). While both benchmarks were sensitive to uncertainty in productivity,  $S_{\text{gen}}$  was more robust (smaller differences in MPE) than  $S_{25\text{th}}$ . For the lower benchmark,  $S_{\text{gen}}$ , Ricker autocorrelation had moderate impacts on performance and the remaining input parameters had relatively weak effects on performance (<50%). For the lower percentile benchmark,  $S_{25\text{th}}$ , initial harvest rates had a moderate effect on benchmark performance, and the remaining input parameters had relatively weak effects on performance (<50%). Similar patterns of results were found for the differences in mean raw error of estimated benchmark from the true value (Appendix, Fig. A1).

Our model assumed spawner abundances at equilibrium,  $S_{\text{eq}}$ , remained constant as productivity varied in sensitivity analyses (as in Holt and Bradford 2011). When we considered an alternate assumption where  $S_{\text{max}}$  remained constant, but  $S_{\text{eq}}$  declined as productivity declined, we found similar results (within 4% MPE). This alternate assumption represents a scenario of simultaneous declines capacity and productivity.

We further explored bivariate sensitivity analyses of the effects of variability in productivity and initial harvest rates on benchmark MPEs. At moderate to high productivity and low initial harvest rates, both  $S_{25\text{th}}$  and  $S_{\text{gen}}$  benchmarks were precautionary (i.e., benchmarks were equal to or higher than “true”  $S_{\text{gen}}$  lower benchmark) (Fig. 10, top left portion of panels). At low productivity and high initial harvest rates (Fig. 10, bottom right corner), neither benchmark is precautionary, but  $S_{\text{gen}}$  performed slightly better (i.e., was slightly closer to the “true” value than  $S_{25\text{th}}$ ).

When we superimposed CU productivities and harvest rates for inner south coast chum salmon, the associated mean percent errors for the estimated lower benchmarks were greater than zero for all CUs (Fig. 10, symbols lie above zero contour line). However, for the 25<sup>th</sup> percentile benchmark, the error bounds crossed the zero contour line for 2 CUs, Georgia Strait and Howe Sound Burrard Inlet, and for the  $S_{\text{gen}}$  benchmark, this was the case for all CUs except Upper Knight. For upper benchmarks, the estimated values were always greater than the true values,

and these positive deviations were greatest for  $S_{75\text{th}}$  benchmark (Fig. 11). The curvilinear relationship between the estimates 80%  $S_{\text{MSY}}$  benchmark and initial harvest rates and productivity were due to the covariance between productivity and carrying capacity when estimating the stock-recruitment relationship and the resulting confounding effect on estimates of  $S_{\text{MSY}}$ .

The global sensitivity analyses showed similar patterns as the univariate and bivariate sensitivity analyses. The mean elemental effects (magnitude of sensitivity) were greatest for productivity for both  $S_{25\text{th}}$  and  $S_{\text{gen}}$  benchmarks (Fig. 12). Initial harvest rates were secondarily important for  $S_{25\text{th}}$  benchmark (Fig. 12b), and the Ricker autocorrelation coefficient and observation errors were also important for  $S_{\text{gen}}$ . The Ricker autocorrelation coefficient was especially important in combination with other input variables for  $S_{\text{gen}}$ , resulting in relatively high standard deviation in elemental effects.

## Objective 2. Adapting benchmarks to changes in productivity

For Objective 2, we simulated an abrupt decline in productivity in the population dynamics sub-model and evaluated benchmarks that used truncated data sets representing either historical baseline conditions (high productivity period) or recent conditions (low productivity period). This data-truncation approach was adopted by Grant et al. (2011) to evaluate biological status of sockeye salmon on the Fraser River using biological benchmarks under the Wild Salmon Policy. Specifically, Grant et al. (2011) found that benchmarks estimated using spawner and recruitment data from only the recent low-productivity period were larger (i.e., more precautionary) than those that were estimated from the entire time-series. This approach has not been applied to other species or regions, where the data required for assessing changes in productivity are often lacking.

### Methods

We adapted the simulation model from Objective 1b in 4 ways:

- The population dynamics sub-model included time-varying productivity modelled as an abrupt decline from  $\log_e(\text{recruits/spawner}) = 2$  to 1 at year 35. The magnitude of these change in productivity was chosen to reflect the magnitude of observed changes in the productivity for chum salmon (Dorner et al. 2008). Evaluating the effects of different temporal different patterns in productivity was outside the scope of this study. We considered both the scenario where  $S_{eq}$  remains constant as productivity changes, and where  $S_{max}$  remains constant, and  $S_{eq}$  declines with productivity. The latter reflects a scenario of changes in productivity and capacity.
- The population model was run over 70 years (35 years prior to and after the productivity shift).
- In the assessment sub-model, benchmark estimation occurred in the final year of the simulation only.
- Assessments used either the entire time-series, the first 30 years (base line, high-productivity period), or the final 30 years (recent, low-productivity period) to estimate benchmarks.

### Results

As expected, regime shifts from high to low productivity were associated with increases in “true” (i.e., deterministic)  $S_{gen}$ , 80% of  $S_{MSY}$ , and  $S_{max}$  when  $S_{eq}$  was assumed constant (Fig. 13a-e). In contrast, when  $S_{max}$  was assumed constant, declines in productivity were associated with declines in “true” 80% of  $S_{MSY}$  and increases in “true”  $S_{gen}$  and  $S_{eq}$  (Fig. 13f-j). The latter assumption incorporates a decline in total capacity of the CU to sustain a population as well as a decline in recruits/spawner at low spawner abundances (see Fig. 3 in Holt and Folkes 2015).

Truncating time-series data used to estimate benchmarks to the recent low-productivity period resulted in lower estimates of productivity (Fig 14a) and higher estimates of  $S_{gen}$  (i.e., more precautionary) (Fig. 14c) under constant  $S_{eq}$ , as expected from previous analyses. The opposite was true when the historical period was used (higher productivity and lower  $S_{gen}$  estimates). Median estimates of  $S_{gen}$  were below the “true” value for all but the most recent estimate (dashed

line Fig. 14c), though the confidence intervals covered the “true” value in all three scenarios. The upper benchmark, 80% of  $S_{MSY}$ , did not change consistently with data truncation (Fig. 14e). Both  $S_{25th}$  and  $S_{75th}$  percentile benchmarks declined (i.e., became less precautionary) when data were truncated to the recent period, but these values were consistently higher than the “true”  $S_{gen}$  and 80% of  $S_{MSY}$  benchmarks, respectively. Although  $S_{gen}$  became more precautionary as data were truncated to recent period and percentile benchmarks became less precautionary, percentile benchmarks were still consistently greater than true values (mean percent errors were  $\gg$  zero) (Fig. 15a)

Similar patterns were observed under the assumption of constant  $S_{max}$  with an abrupt decline in productivity and  $S_{eq}$  (Fig. 15b & 16), with two exceptions. When only the recent data were used, the estimate of productivity did not decline compared with using the entire data set (Fig. 16a), though the estimate of  $S_{eq}$  did decline (Fig. 16b). Confounding between estimates of productivity and  $S_{eq}$  results in a relatively low (instead of high) value for  $S_{gen}$  when only recent data are used (Fig. 16c).

## Discussion and Conclusions

Our retrospective analysis of chum CUs on inner south coast of BC indicates that 25<sup>th</sup> and 75<sup>th</sup> percentile benchmarks are generally higher than stock-recruitment based benchmarks adopted under the Wild Salmon Policy, and are therefore a precautionary choice in data-limited situations. The few exceptions in our analyses did not occur in the most recent year (i.e., occurred in retrospective assessments that used shorter time-series) and were associated with either long periods of very low escapement or large, abrupt increases in productivity or escapement. Upper percentile benchmarks ( $S_{75th}$ ) were considerably higher than upper stock-recruitment based benchmarks ( $>>80\% S_{MSY}$ ), and by definition, green status occurred in only a quarter of observed years. If upper percentile benchmarks are used to inform fisheries management targets, these will likely lie above  $S_{MSY}$  levels resulting in harvests below MSY.

For the seven CUs analyzed here, benchmarks derived from hierarchical Ricker models were virtually indistinguishable from those estimated using standard Ricker models. In the retrospective analysis, the standard Ricker model and hierarchical Ricker model gave the same status for 99% of CU-year combinations. However, benchmarks derived from the hierarchical Ricker model were generally less uncertain than those from the standard model. Given large uncertainties in stock-recruitment data and inconsistent time-series for chum salmon in BC, a hierarchical approach is recommended over standard Ricker model when there is support for the assumption of similar productivities among CUs.

Our retrospective analysis illustrates that most inner south coast chum CUs have shown considerable change in conservation status over time and these changes depend on the benchmarks used. Percentile-based benchmarks were associated with higher variability in status over time than stock-recruitment benchmarks. One caveat on the application of percentile-based benchmarks is that uncertainties in benchmarks are not provided. Nevertheless, status determined using percentile-based benchmarks were consistently more precautionary than Ricker-based benchmarks, and are therefore a viable choice for use in sites where Ricker-based benchmarks cannot be calculated or are highly uncertain.

Similarly, our simulation model suggests that the lower percentile benchmark,  $S_{25th}$ , tends to be more precautionary than the corresponding stock-recruitment based benchmark,  $S_{gen}$ , when historical harvest rates are moderate to low, current harvest rates are low, and productivity is moderate to low. At low productivity and high exploitation rates, neither benchmark is precautionary, but the percentile benchmark is especially negatively biased. Although most CUs in our study had historical exploitation rates and productivities associated with relatively precautionary estimates of benchmarks, uncertainties in the remaining input parameters may affect benchmark performance. Specifically,  $S_{gen}$  performance is also sensitive to Ricker autocorrelation coefficient, Ricker sigma (residual variance), and observation errors in abundances.

Our retrospective analyses assumed constant productivity over time, but observed temporal variability in productivity (measured in Ricker  $\alpha$ ) for south coast chum salmon CUs suggests that the Ricker stock-recruit relationships may vary through time (here and documented in Malick and Cox 2016). Indeed, there is widespread evidence for abrupt regime shifts in salmon productivity followed by relatively constant periods (Beamish et al. 1999, Hare et al. 1999). Given widespread changes in productivity, a data-truncation approach that uses data from recent



period or historical base-line period of relatively consistent productivity may result in benchmarks that are more precise and in some cases more precautionary. Although truncation of time-series data to the most recent time period has been suggested as a method to account for declines in productivity in a precautionary manner (Grant et al. 2011), this approach may result in a “shifting baseline” for percentile-based benchmarks since they tend to decline with abundances. In other words, declines in productivity affect  $S_{\text{gen}}$  and percentile-based benchmarks in opposite directions.

Our simulation model suggests that truncating data to a recent low-productivity period only results in  $S_{\text{gen}}$  values that are more precautionary and percentile-based benchmarks that are less precautionary under the scenario of constant  $S_{\text{eq}}$ . When  $S_{\text{eq}}$  also declines with productivity (simultaneous decline in production), then both  $S_{\text{gen}}$  and  $S_{25\text{th}}$  benchmarks become less precautionary, but this change is greatest for  $S_{25\text{th}}$  benchmark. Despite these divergent trends with data truncation, percentile-based benchmarks remain more precautionary overall (with caveats from Objective 1 of this study described above).

Any changes to benchmarks in response to changes in productivity (i.e., data truncation) will require careful consideration of strength of evidence, causal mechanisms, and reversibility of changes, among other factors, and should be implemented on a case-by-case basis. If the observed decline in abundances is due to density-independent changes in productivity that are reversible, then precautionary (i.e., relatively high) benchmarks are warranted to maintain resilience of the CU. If the observed decline is due to density-dependent changes in capacity that are well understood and irreversible (i.e., a persistent shift to a low-production regime), reductions in benchmarks may be warranted in rare cases to reflect this decline in production. Duplisea and Cadigan (2012) provide recommendations on the conditions that would be required to make such adjustments.

One limitation in our data truncation analysis is that it assumes that shifts in productivity are detected accurately, and data are truncated to within a specific regime. The results of our Ricker model with time-varying productivity in Objective 1a demonstrate that temporal estimates of productivity are highly uncertainty, making regime shifts challenging to identify. Developing improved methods for detecting regime shifts within these data sets is an area for future investigation.

## References

- Beamish R, Noakes D, McFarlane G, Klyashtorin L, Ivanov V, Kurashov V. 1999. The regime concept and natural trends in the production of Pacific salmon. *Canadian Journal of Fisheries and Aquatic Sciences* 56:516-526.
- Clark RA, Eggers DM, Munro AR, Fleischman SJ, Bue BG, Hasbrouck JJ. 2014. An Evaluation of the Percentile Approach for Establishing Sustainable Escapement Goals in lieu of Stock Productivity Information. Fishery Manuscript No. 14-06, Alaska Department of Fish and Game.
- Collie JS, Peterman RM, Zuehlke BM, Chen Y. 2012. A fisheries risk-assessment framework to evaluate trade-offs among management options in the presence of time-varying productivity. *Canadian Journal of Fisheries and Aquatic Sciences* 69:209-223.

Cousens NBF. 1982. A review of salmon escapement estimation techniques. Can. Tech. Rep. Fish. Aquat. Sci. 1982/1108.

DFO (Fisheries and Oceans Canada). 2005. Canada's Policy For Conservation of Wild Pacific Salmon. Cat. No. Fs23-476/2005E, DFO (Fisheries and Oceans Canada).

Dobson D, O'Brien D, Pestal G, Fisheries Do, Oceans V, BC. Fisheries, Branch AM. 2009. Certification Unit Profile: West Coast Vancouver Island Chum Salmon. Fisheries & Aquaculture Management Branch, Department of Fisheries and Oceans.

Dorner B, Peterman RM, Haeseker SL. 2008. Historical trends in productivity of 120 Pacific pink, chum, and sockeye salmon stocks reconstructed by using a Kalman filter. Canadian Journal of Fisheries and Aquatic Science 65: 1842–1866

Duplisea D, Cadigan N. 2012. Proceedings of the national workshop for technical expertise in stock assessment (TESA): maximum sustainable yield (MSY) reference points and the precautionary approach when productivity varies. Can. Sci. Adv. Sec. Proc. Ser. 2012/055.

Fair L, Willette T, Erickson J, Yanusz R, McKinley T. 2010. Review of Salmon Escapement Goals in Upper Cook Inlet, Alaska, 2011. Alaska Department of Fish and Game. Fishery Manuscript Series No. 10-06.

Fleischman SJ, Catalano MJ, Clark RA, Bernard DR, Chen Y. 2013. An age-structured state-space stock–recruit model for Pacific salmon (*Oncorhynchus* spp.). Canadian Journal of Fisheries and Aquatic Sciences 70:401-414.

Grant SCH, MacDonald BL, Cone TE, Holt CA, Cass A, Porszt EJ, Hume JMB, Pon LB. 2011. Evaluation of Uncertainty in Fraser Sockeye (*Oncorhynchus nerka*) Wild Salmon Policy Status using Abundance and Trends in Abundance Metrics. DFO Can. Sci. Advis. Sec. Res. Doc. 2011/087:viii + 183 p.

Hare SR, Mantua NJ, Francis RC. 1999. Inverse Production Regimes: Alaska and West Coast Pacific Salmon. Fisheries 24:6-14.

Hilborn R, Schmidt D, English K, Devitt S. 2013. British Columbia Chum Salmon (*Oncorhynchus keta*) Fisheries; British Columbia Coastal and Adjacent Canadian Pacific EEZ Waters. Ref: (82506), Intertek Moody marine.

Hilborn R, Walters CJE. 1992. Quantitative Fisheries Stock Assessment; Choice Dynamics & Uncertainty. Springer US.

Holt C, Cass A, Holtby B, Riddell B. 2009. Indicators of status and benchmarks for conservation units in Canada's Wild Salmon Policy. DFO Can. Sci. Advis. Sec. Res. Doc. 2009/058.

Holt CA, Bradford MJ. 2011. Evaluating benchmarks of population status for Pacific salmon. North American Journal of Fisheries Management 31:363-378.

Holt CA, Folkes MJP. 2015. Cautions on using percentile-based benchmarks of status for data-limited populations of Pacific salmon under persistent trends in productivity and uncertain outcomes from harvest management. *Fisheries Research* 171:188-200.

Holt CA, Irvine JR. 2013. Distinguishing benchmarks of biological status from management reference points: A case study on Pacific salmon in Canada. *Environmental Conservation* 40:345-355.

Korman J, Cox-Rogers S, English K. 2013. Benchmark Analysis for Pacific Salmon Conservation Units in the Skeena Watershed. Pacific Salmon Foundation.

Korman J, Peterman RM, Walters CJ. 1995. Empirical and theoretical analyses of correction of time-series bias in stock-recruitment relationships of sockeye salmon (*Oncorhynchus nerka*). *Canadian Journal of Fisheries and Aquatic Sciences* 52:2174-2189.

Malick MJ, Cox SP. 2016. Regional-Scale Declines in Productivity of Pink and Chum Salmon Stocks in Western North America. *PLoS One* 11:e0146009.

McElhany P, Ruckelshaus M, Ford M, Wainwright T, Bjorkstedt E. 2000. Viable Salmonid Populations and the Recovery of Evolutionary Significant Units. NOAA Technical Memorandum NMFS-NWFSC-42, U.S. Dept. Commer.

Morris DJ, Speirs DC, Cameron AI, Heath MR. 2014. Global sensitivity analysis of an end-to-end marine ecosystem model of the North Sea: Factors affecting the biomass of fish and benthos. *Ecological Modelling* 273:251-263.

Pacific Salmon Commission Joint Chum Technical Committee 2013. Joint Chum Technical Committee 2010 Summary Report. TCCHUM (15)-1.

Peterman RM, Pyper BJ, MacGregor BW. 2003. Use of the Kalman filter to reconstruct historical trends in productivity of Bristol Bay sockeye salmon (*Oncorhynchus nerka*). *Canadian Journal of Fisheries and Aquatic Sciences* 60:809-824.

Plummer M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing. Vienna, Austria.

Pujol G, et al. 2015. sensitivity: Sensitivity Analysis: R package version 1.11.1.

Pyper BJ, Mueter FJ, Peterman RM, Blackburn DJ, Wood CC. 2002. Spatial Covariation in Survival Rates of Northeast Pacific Chum Salmon. *Transactions of the American Fisheries Society* 131:343-363.

R Development Core Team. 2016. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ricker WE. 1975. Computation and Interpretation of Biological Statistics of Fish Populations. Bulletin of the Fisheries Research Board of Canada, No 119., Ottawa.

Su Y-S, Yajima M. 2012. R2jags: A Package for Running jags from R. R package version 0.03-08, URL <http://CRAN.R-project.org/package=R2jags>.

Szerlong RG, Rundio DE. 2008. A statistical modeling method for estimating mortality and abundance of spawning salmon from a time series of counts. Canadian Journal of Fisheries and Aquatic Sciences 65:17-26.

Van Will PR. 2014. Inner South Coast Chum Stock Reconstructions. Report submitted to the Marine Stewardship Council.

Walters CJ, Martell SJD. 2004. Fisheries ecology and management. Princeton University Press.

**Figures:**

Fig. 1. Prior distribution of Ricker  $\alpha$  parameter (productivity).

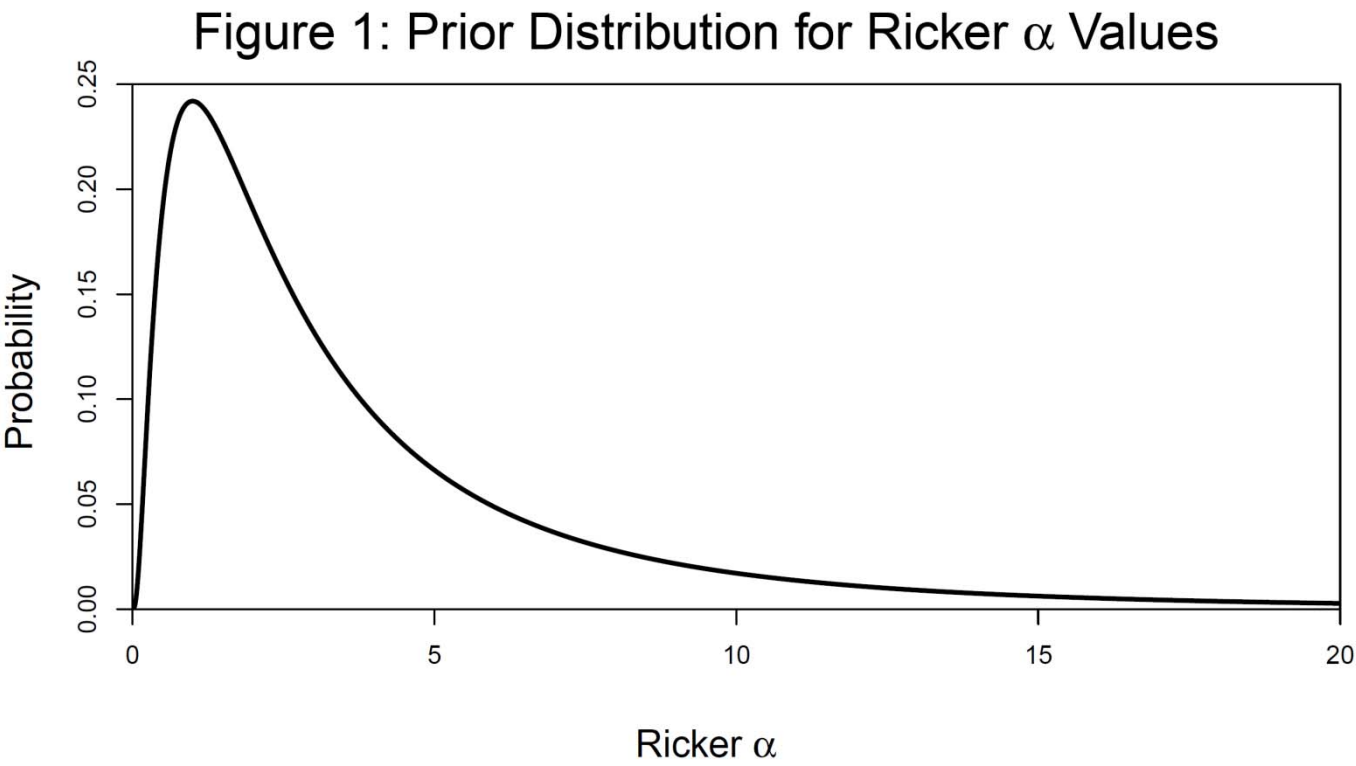


Fig. 2. Model estimates for Ricker  $\alpha$  (top row) and  $S_{max}$  (bottom row) across prior distributions (uniform and lognormal) for  $S_{max}$  and standard and hierarchical Ricker model structures (solid and dotted lines, respectively) for each CU in 2012. Circles indicate posterior medians, and lines indicate 95% credible intervals of estimates.

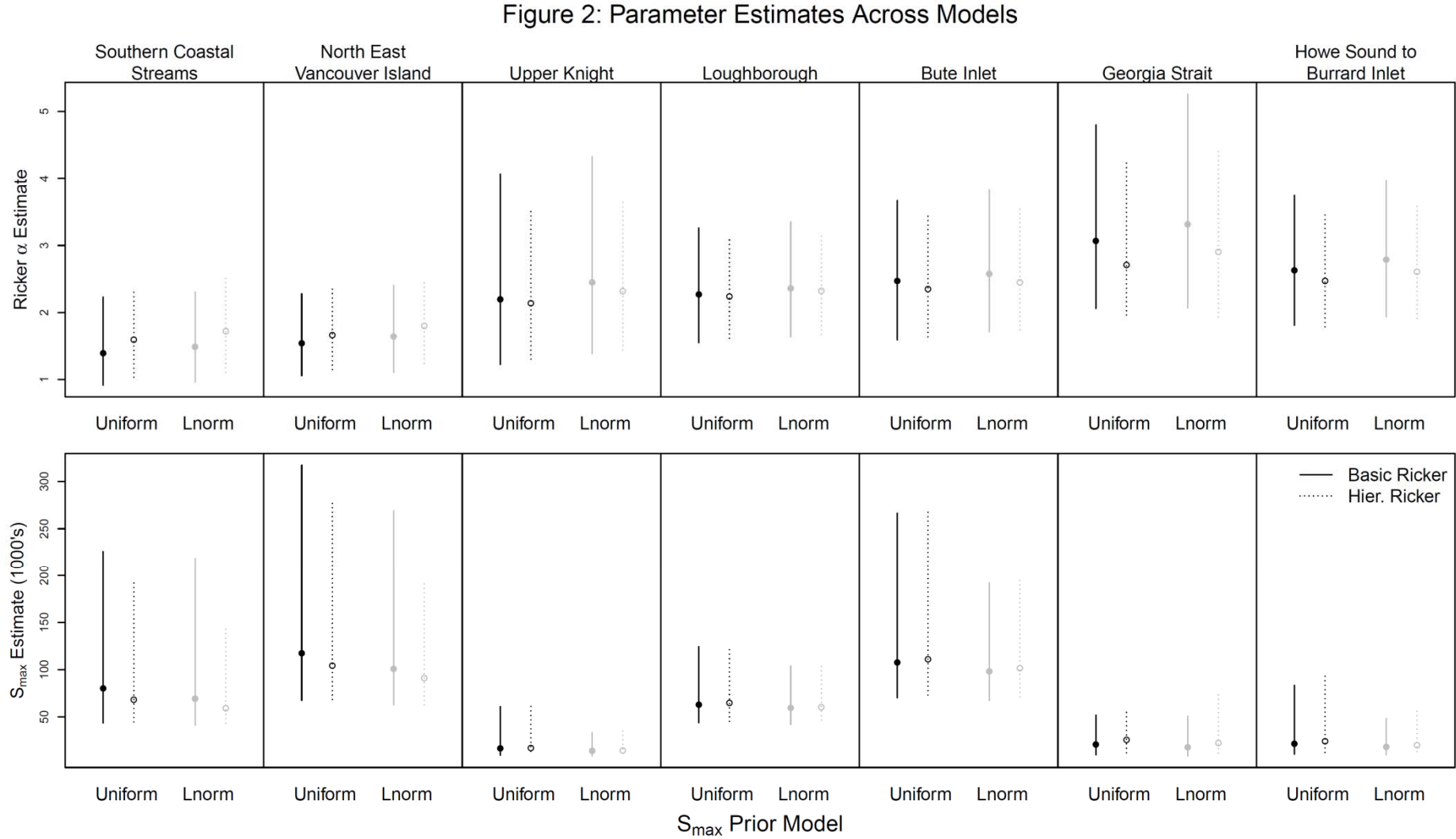


Fig. 3. Observed spawner-recruit data over time, with fitted Ricker curves and associated benchmarks for (i) the standard Bayesian Ricker model, and (ii) the Bayesian hierarchical Ricker model. Shaded regions indicate 95% credible intervals, delineated by 2.5<sup>th</sup> and 97.5<sup>th</sup> posterior densities. Red and green circles on *x*-axis identify percentile-based benchmarks ( $S_{25th}$  and  $S_{75th}$ , respectively). Cross indicates most recent data point, for brood year 2006. Colours of points increase in darkness as years progress towards the current year.

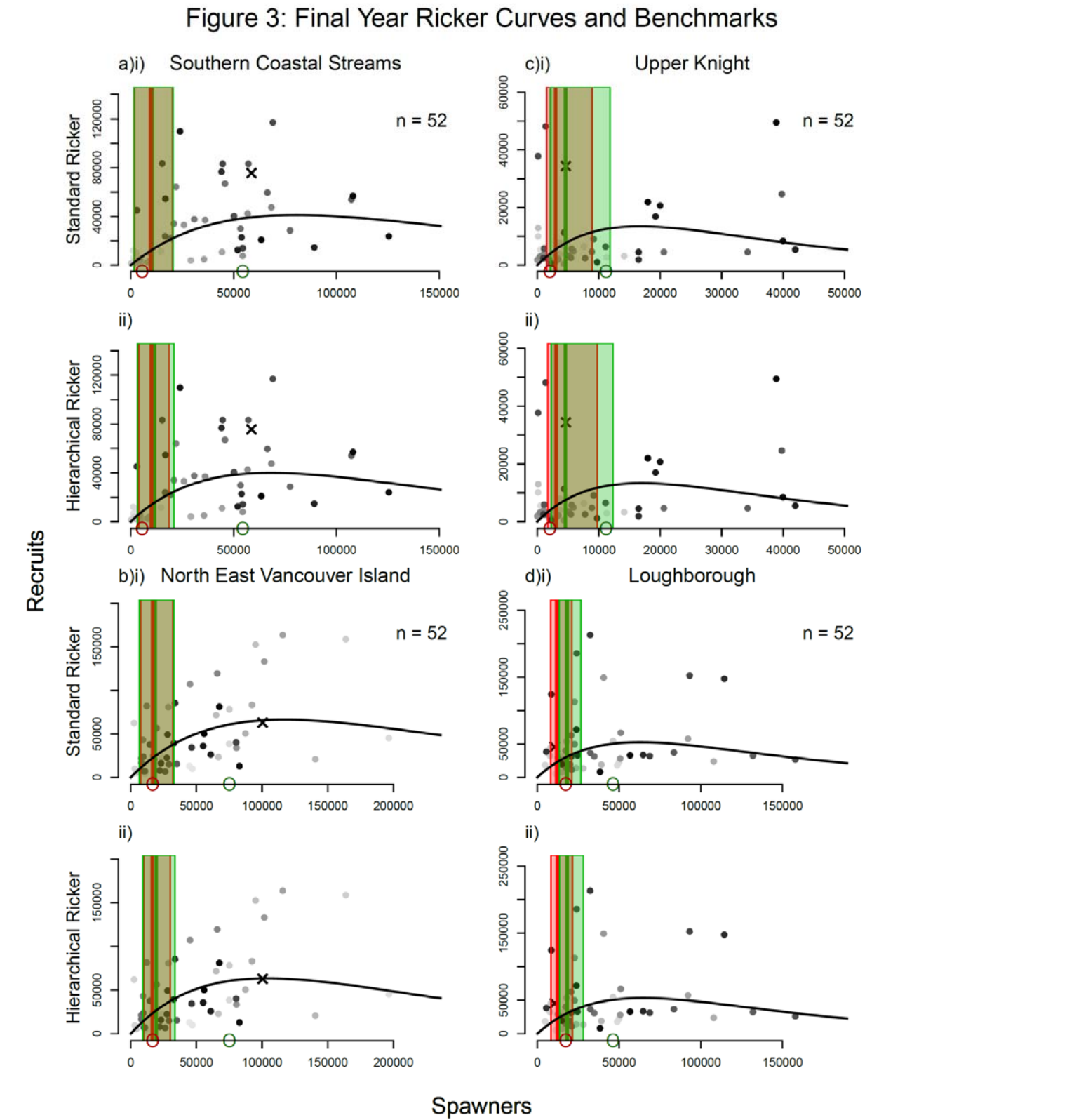


Figure 3: Final Year Ricker Curves and Benchmarks

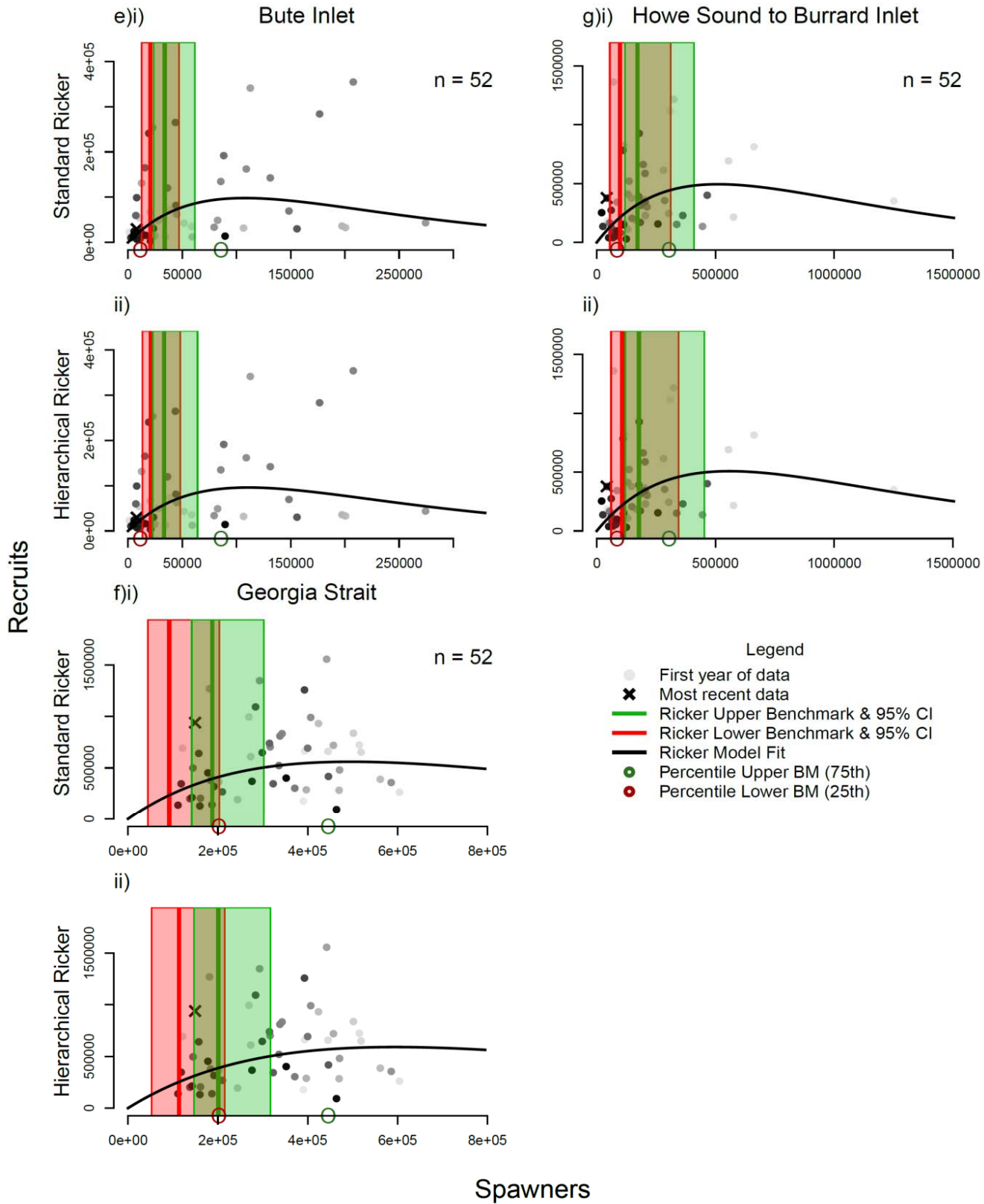




Fig. 4. Raw and generational average escapement over time, with retrospective conservation benchmarks overlaid for three benchmarks types: (i) standard Ricker model; (ii) hierarchical Ricker model; and (iii) percentile. Shaded regions indicate 95% credible intervals, delineated by 2.5<sup>th</sup> and 97.5<sup>th</sup> posterior densities. Retrospective benchmarks use all available data up to that year to estimate benchmark values.

Figure 4: Conservation Benchmarks Over Time

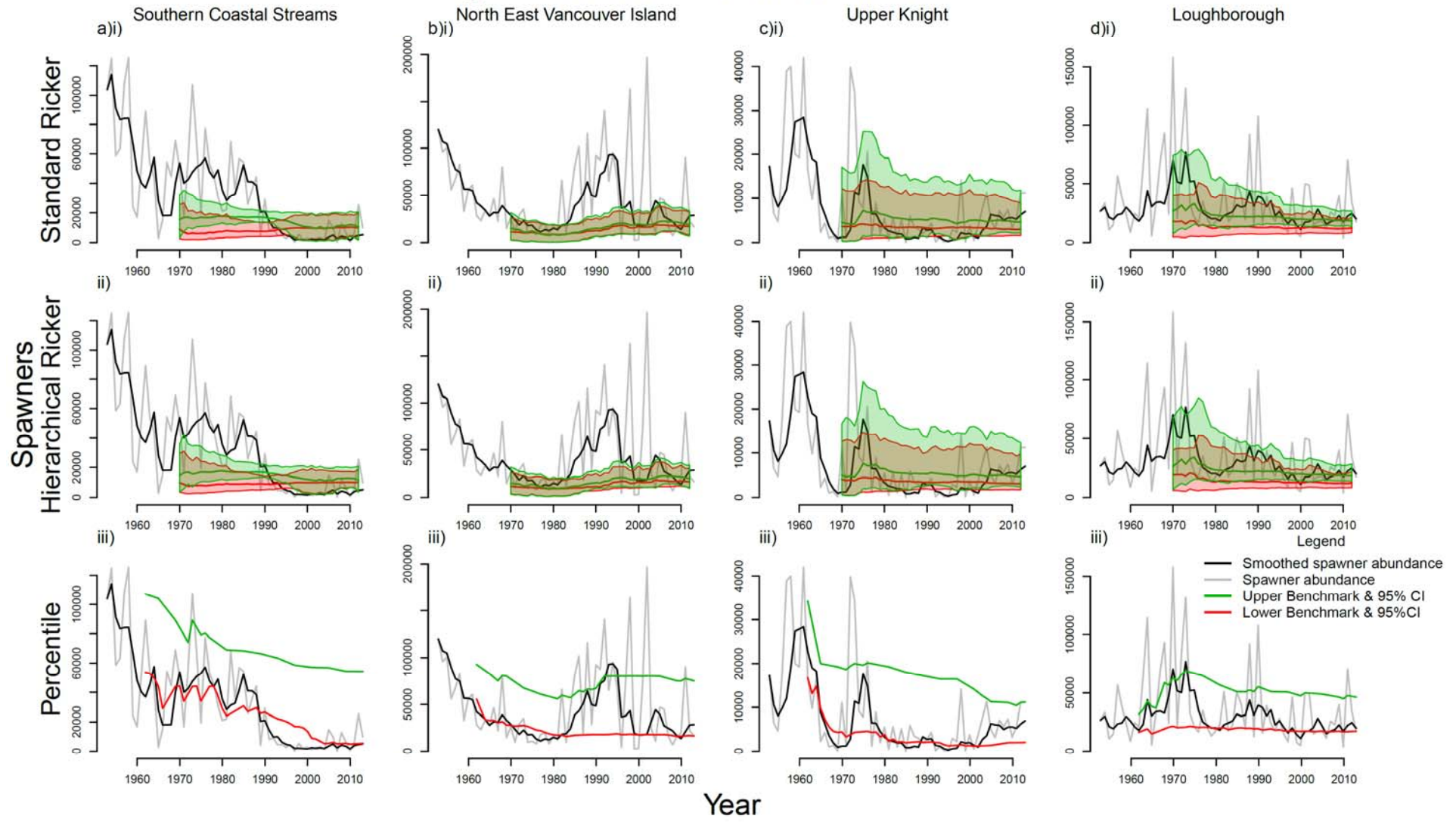


Figure 4: Conservation Benchmarks Over Time

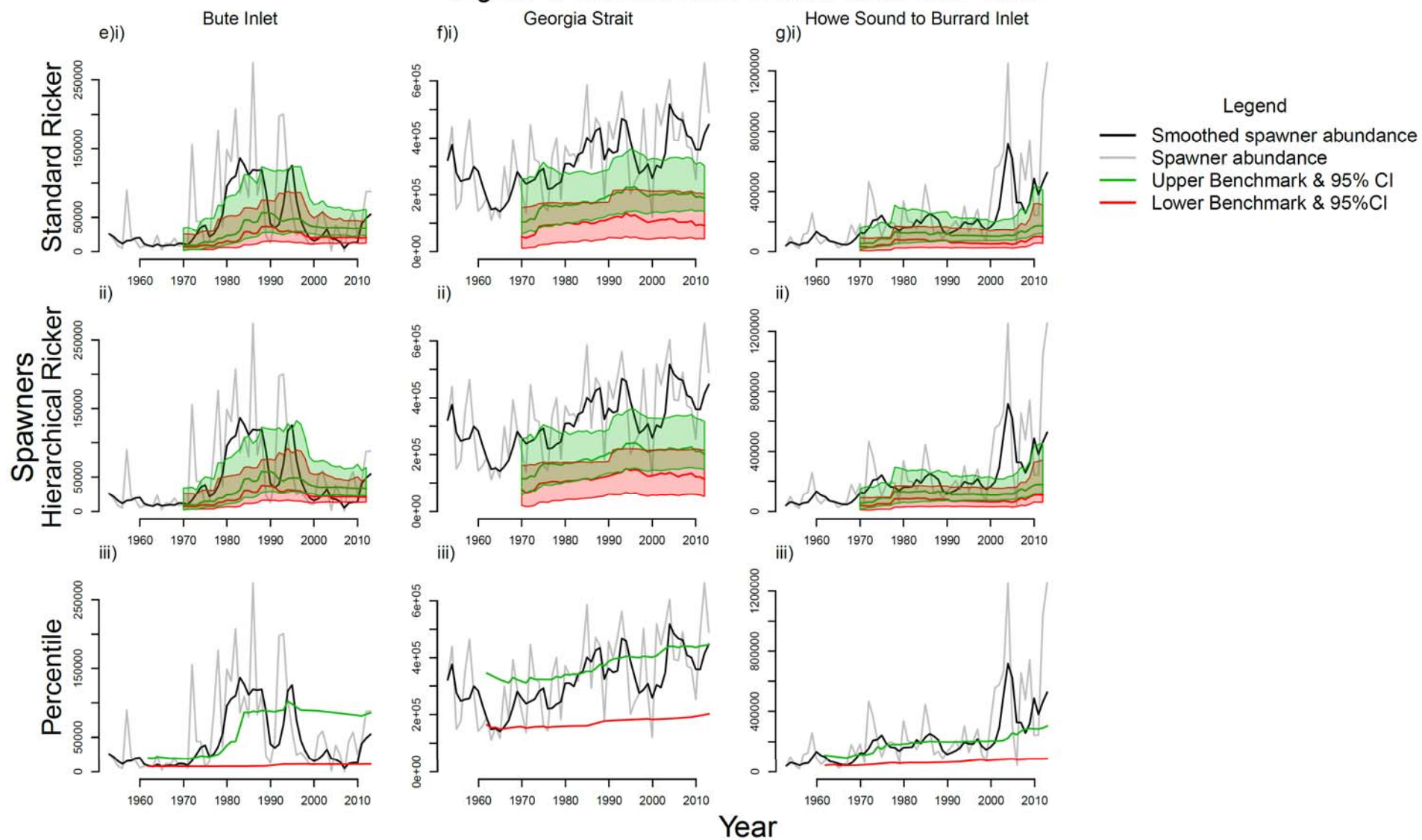


Fig. 5. Standardized raw and generational average escapements across CUs, with conservation status indicated by coloured bars below. Transparent bars indicate upper and lower credible interval bounds, based on 2.5<sup>th</sup> and 97.5<sup>th</sup> posterior densities of estimated parameters. Gaps exist for Southern Coastal Streams and North East Vancouver Island because status was not assessed when  $\alpha$  values were < 1.5, as suggested by Holt and Ogden (2013).

Figure 5: Status Over Time

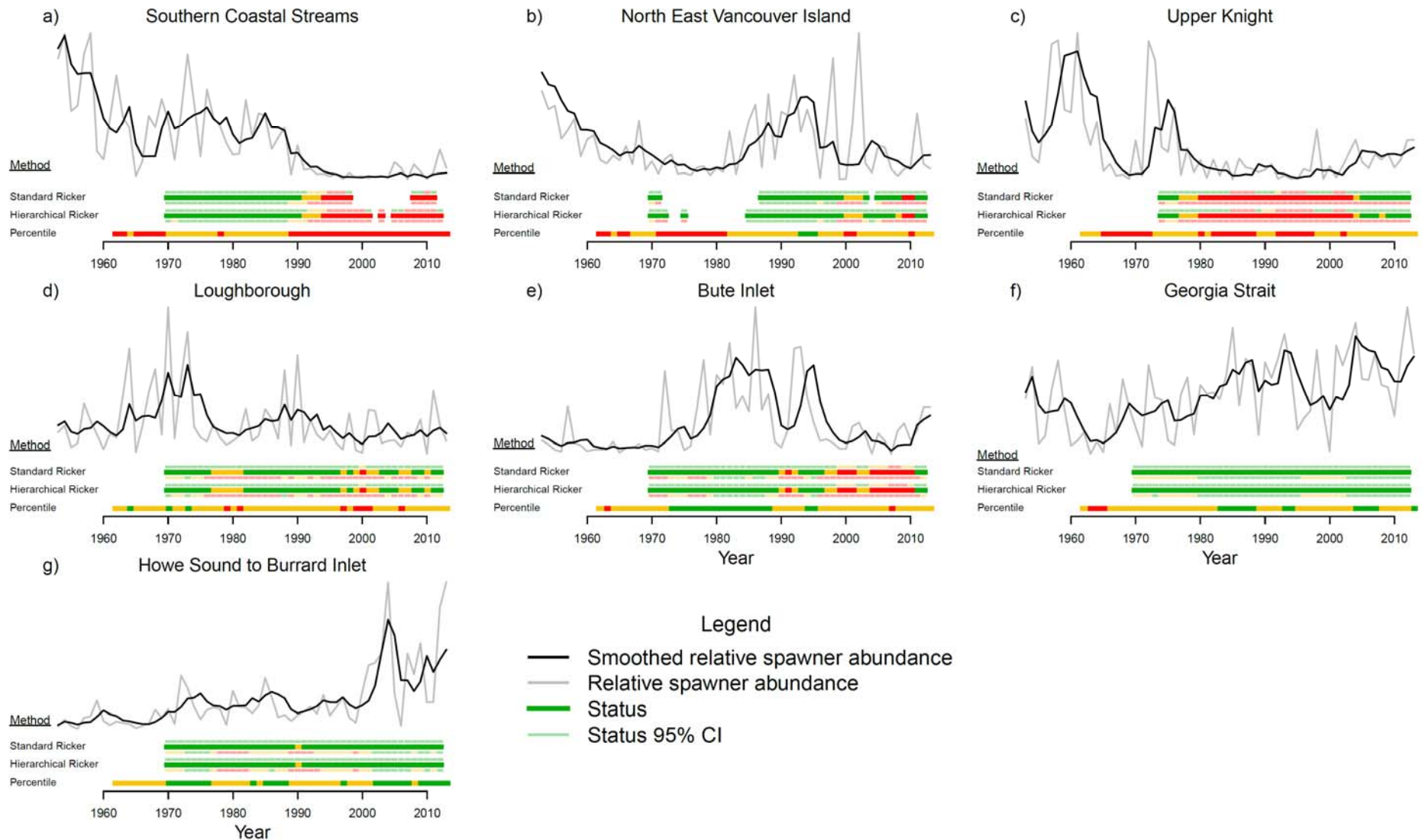


Fig. 6. Estimated Ricker  $\alpha$  values using a recursive Bayes model, which allows  $\alpha$  to vary over time, within a given CU. Grey shaded polygons indicate 95% credible intervals based on posterior densities of estimated  $\alpha$  values.

Figure 6: Ricker  $\alpha$  Estimated by Recursive Bayes

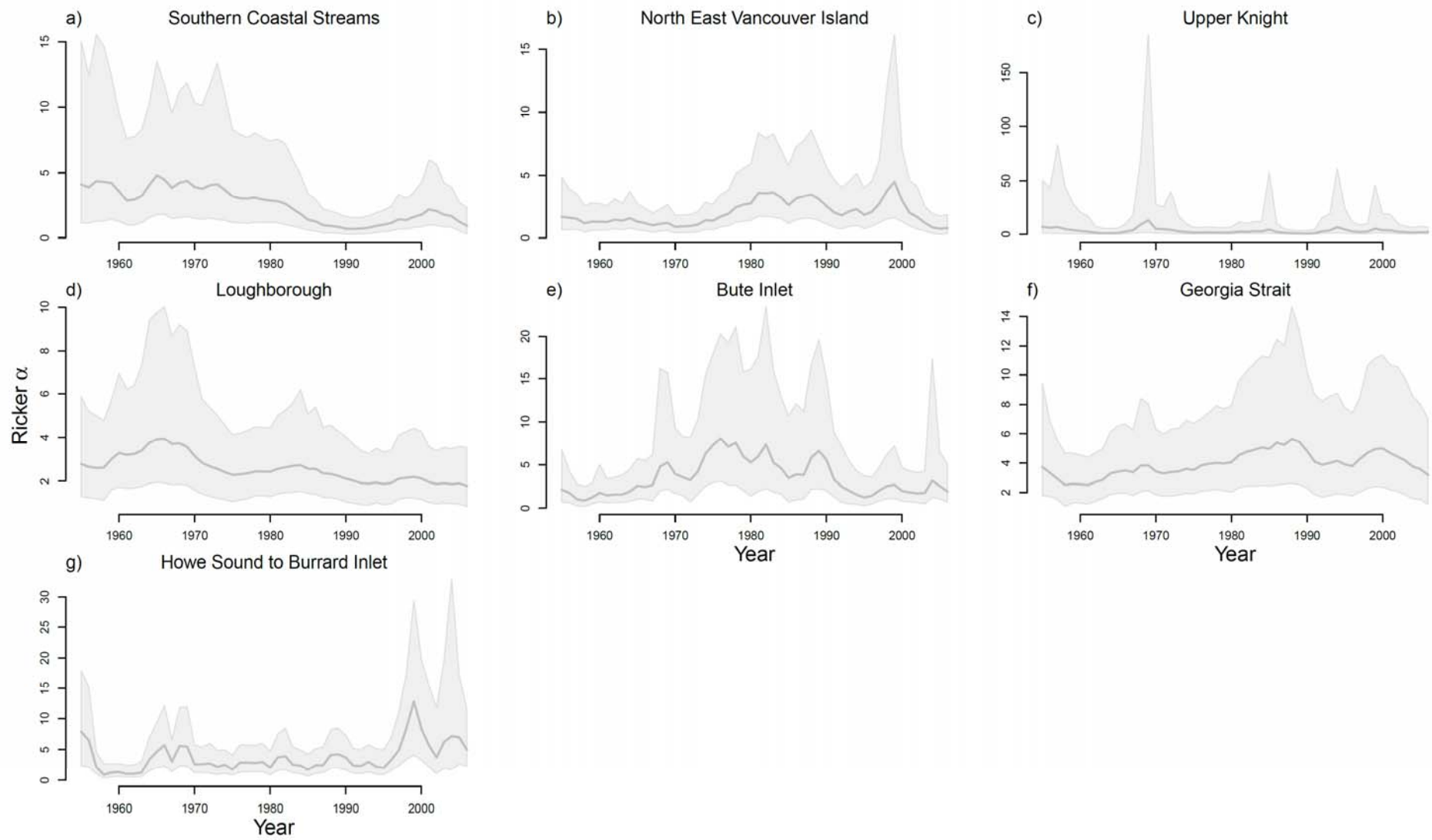


Fig. 7. Schematic of simulation model used to evaluate benchmark performance.

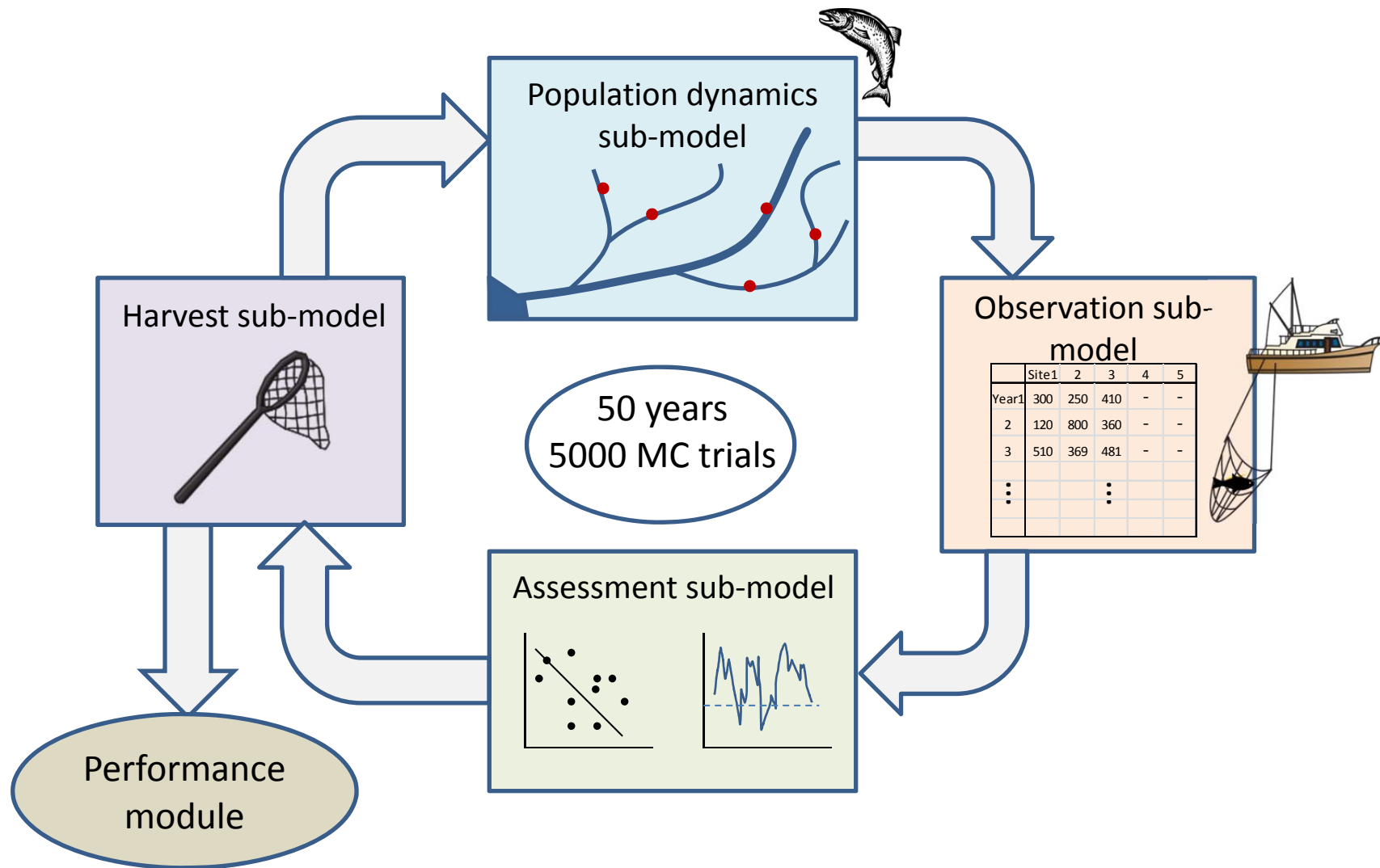




Fig 8(a). Time-series of observed spawner abundances (black line) and “true” spawner abundances (grey line) and benchmarks for one Monte Carlo trial. Benchmarks are estimated annually base on all data up until that year: annual estimates of 80%  $S_{MSY}$  (upper benchmark, green dashed line), annual estimates of  $S_{gen}$  (lower benchmark, red dashed line), the 75<sup>th</sup> percentile benchmark (green dotted line), and the 25<sup>th</sup> percentile (red dotted line). 95% confidence intervals are shown for estimates of stock-recruitment based benchmarks (green and red shading for the upper and lower benchmarks, respectively). Vertical dashed line indicates the end of the 20-year initialization period. (b) Observed spawner and recruitment data (solid black dots) and “true” data (grey hollow dots) for the final year of one Monte Carlo trial. The “true” underlying stock-recruitment relationship is shown with the grey curve and the estimate based on observed data is shown with the black curve. (c) Mean percent error between estimated and “true” benchmark averaged over all Monte Carlo trials. Red bars are the mean percent error from the “true”  $S_{gen}$  (lower benchmark), and green bars are the mean percent error from the “true” 80%  $S_{MSY}$ . \* is mean percent error =229% for the  $S_{75th}$  benchmark.

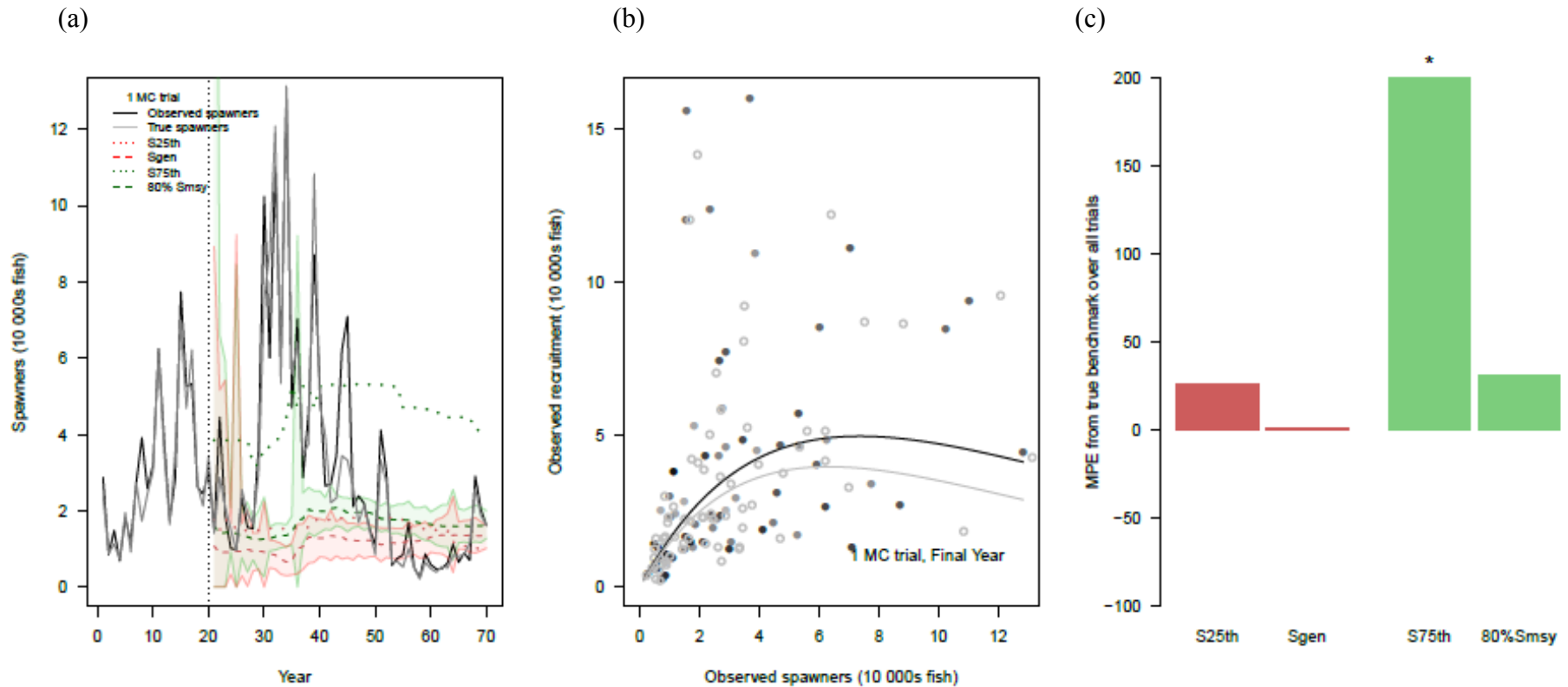


Fig. 9. Difference in the mean percent error, MPE, of estimated lower benchmark ( $S_{\text{gen}}$ , (a), and  $S_{25\text{th}}$  (b)) and the “true” lower benchmarks (“true”  $S_{\text{gen}}$ ), between sensitivity analyses listed on the  $x$ -axis and the base case scenario. Black bars are analyses where the input parameter was increased relative to the base case (see Table 1); white bars are analyses where the input parameter was reduced relative to the base case. Positive values indicate sensitivity analyses where MPE increased under that change in input parameter from the base case; negative values indicate analyses where the MPE declined under that change in input parameter. Asterisks denote values higher than the limit of the  $y$ -axis: 144% (a) and 780% (b).

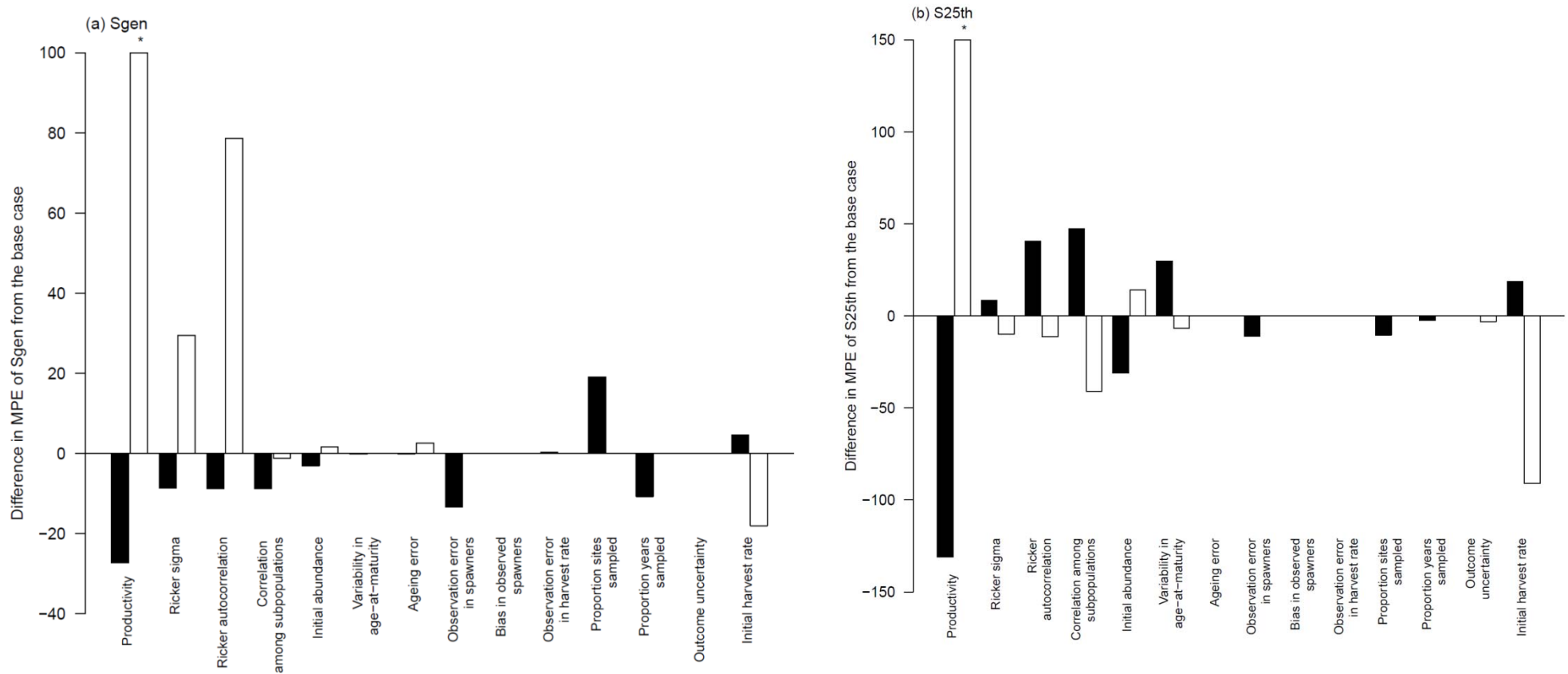


Fig. 10. Mean percent error, MPE, of the estimated lower benchmark ( $S_{25th}$  (a), and  $S_{gen}$  (b)) from the “true”  $S_{gen}$  value along a gradient in initial harvest rates ( $x$ -axis) and productivities ( $y$ -axis) derived from a simulation model of a hypothetical salmon CU. Symbols indicate MPE of CUs assuming productivities estimated from hierarchical Ricker models and mean harvest rates over available time-series for each CU. Y-error bars represent the 95% credible intervals of the estimate of productivity. X-error bars are the standard deviation of historical harvest rates. SCS is Southern Coastal Streams, NEVI is North East Vancouver Island, UK is Upper Knight, LB is Loughborough, GS is Georgia Strait, and HSBI is Howe Sound/Burrard Inlet.

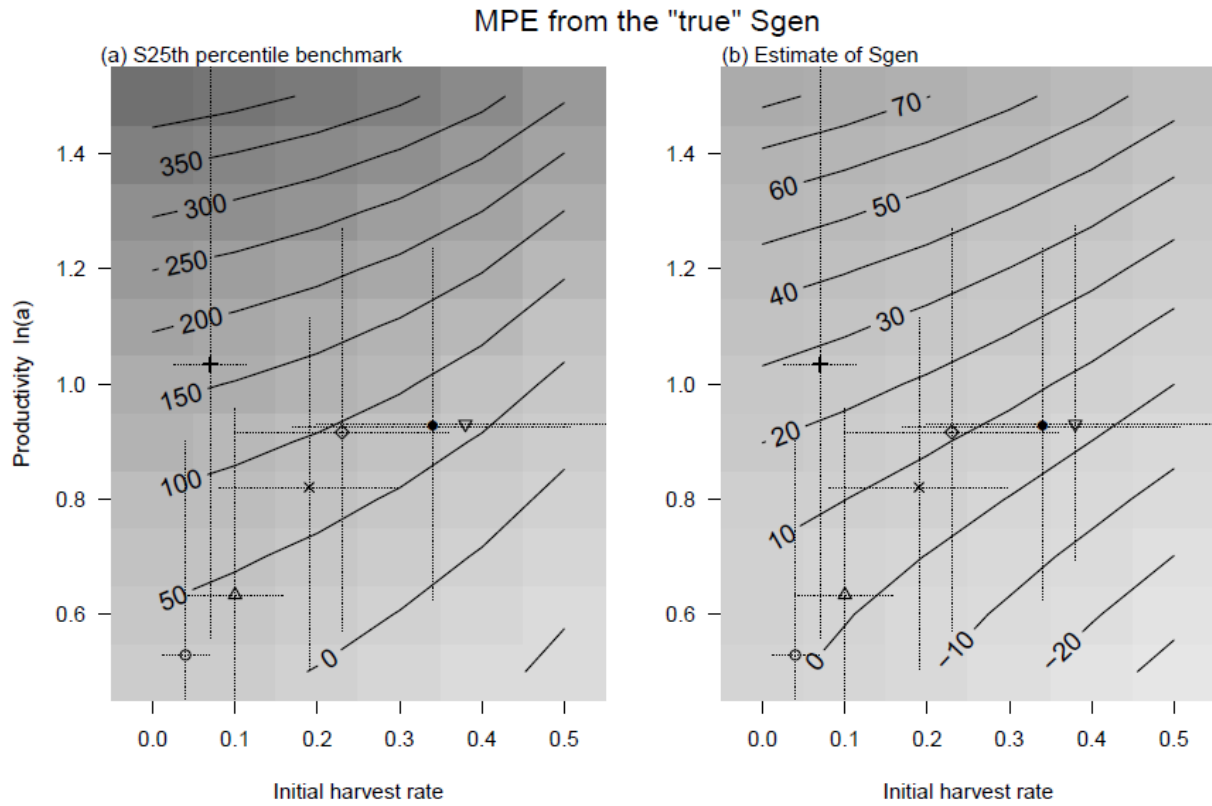




Fig. 11. Mean percent error of the estimated upper benchmark ( $S_{75\text{th}}$  percentile (a), and 80%  $S_{\text{MSY}}$  (b)) from the “true” 80%  $S_{\text{MSY}}$  value along a gradient in initial harvest rates ( $x$ -axis) and productivities ( $y$ -axis) derived from a simulation model of a hypothetical salmon CU. Symbols indicate MPE of CUs assuming productivities estimated from hierarchical Ricker models and mean harvest rates over available time-series for each CU. Y-error bars represent the 95% credible intervals of the estimate of productivity. X-error bars are the standard deviation of historical harvest rates. See caption to Fig. 10 for CU definitions.

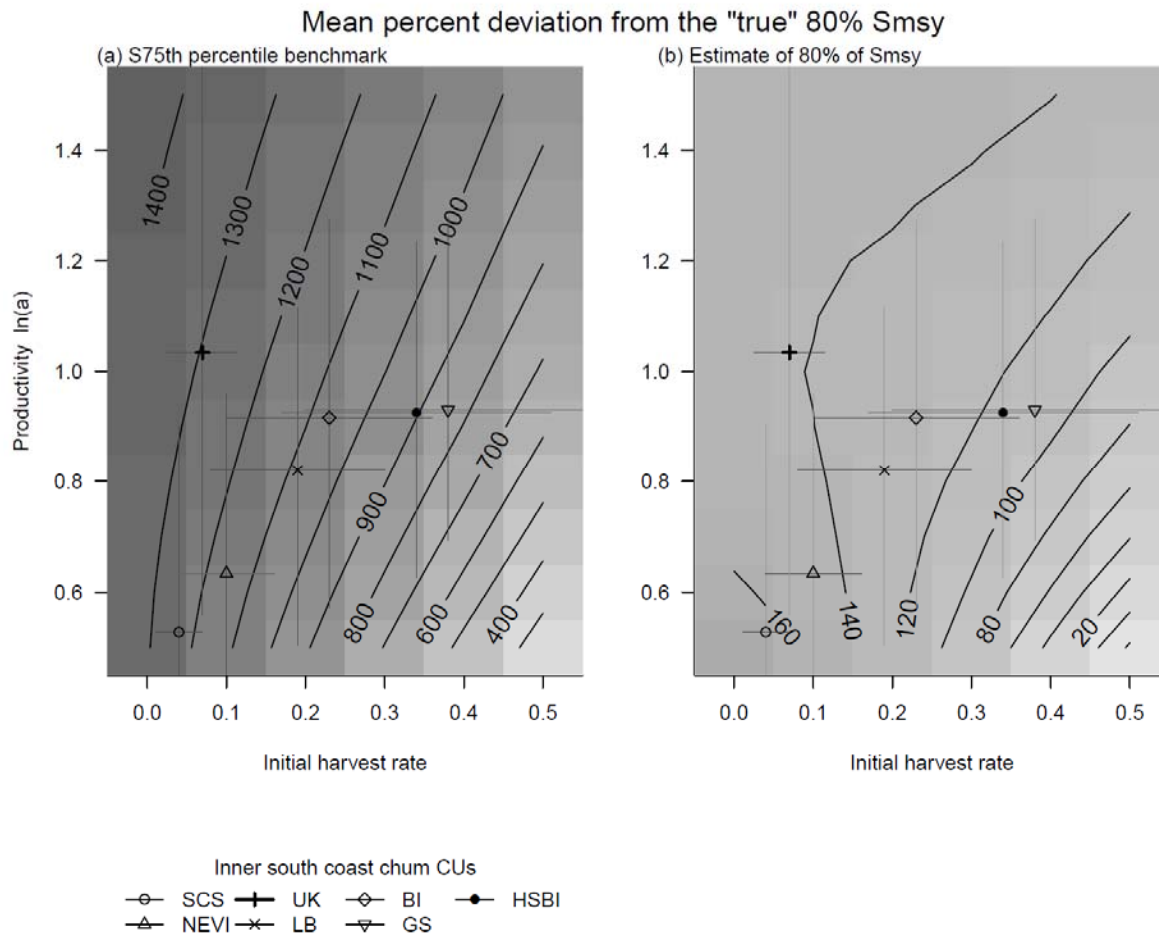


Fig. 12. Sensitivity indices of the effects of individual variables (mean elemental effect, x-axis) and interactions among variables (standard deviation in elemental effects, y-axis). Indices were derived from the Morris method, a global sensitivity analyses for the mean percent error of estimated lower benchmarks ( $S_{\text{gen}}$  (a), and  $S_{25\text{th}}$  (b)) from “true” benchmarks. Input variables with values >100 on either axis are labelled.

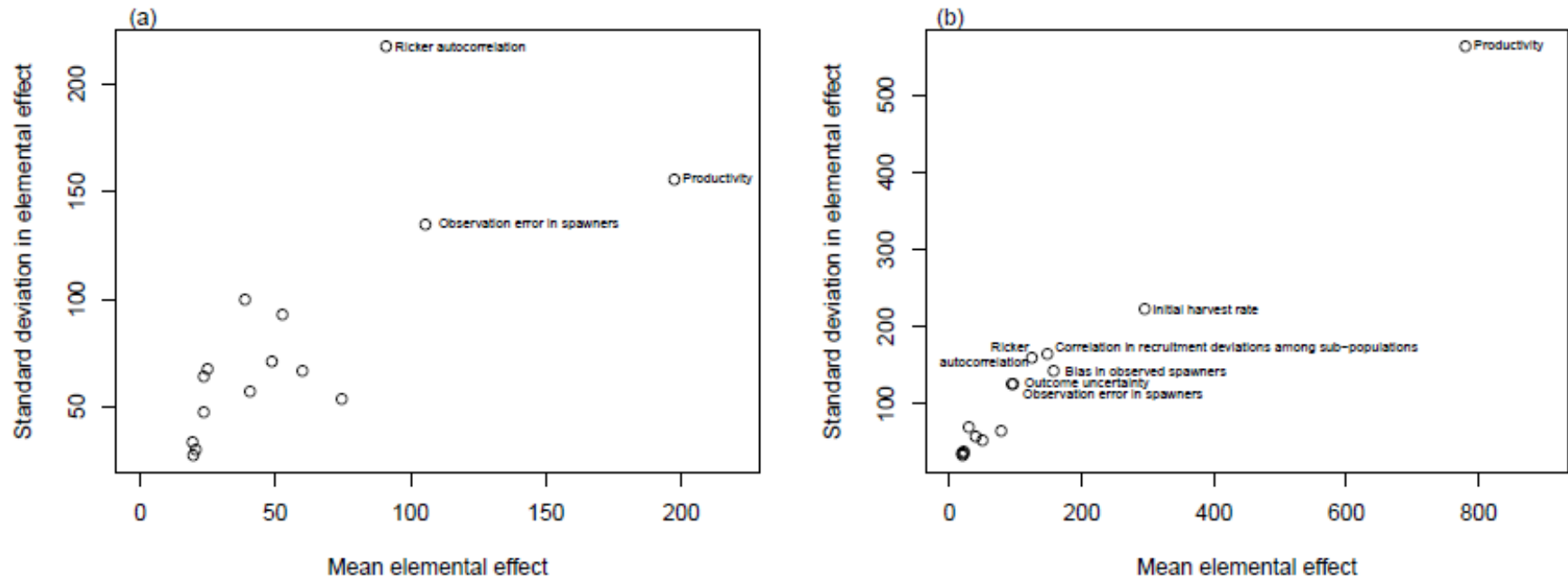


Fig. 13. “True” values of Ricker parameters and benchmarks under assumption of constant spawners at equilibrium,  $S_{eq}$ , (a-e) or constant spawner abundances at maximum recruitment  $S_{max}$  (f-j), with abrupt changes in productivity in year 35 of simulation.

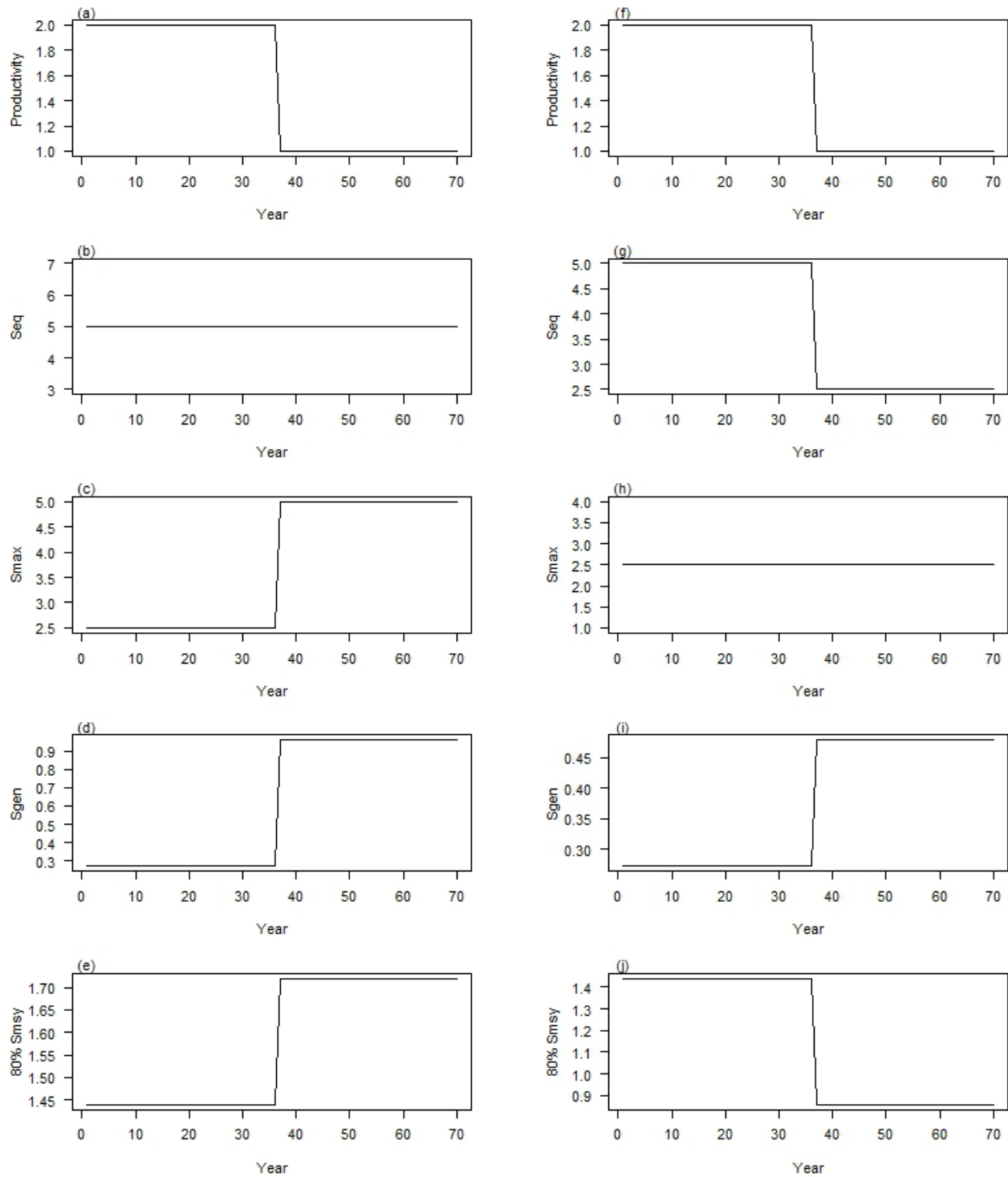


Fig. 14. Box plots of parameter and benchmark values in final year of the simulation averaged over all Monte Carlo trials, using either the first 30 years of data (dark grey boxes), all 70 years of data (light grey boxes), or the most recent 30 years of data (white boxes). Boxes represent the lower quartiles, medians, and upper quartiles of the parameter distribution. Whiskers are the 95% confidence intervals. Dashed lines represent the “true” value for each parameter. For the percentile benchmarks, the dashed lines represent the “true”  $S_{\text{gen}}$  (e) and “true” 80% of  $S_{\text{MSY}}$  (f).  $S_{\text{eq}}$  was held constant in simulations as productivity varied.

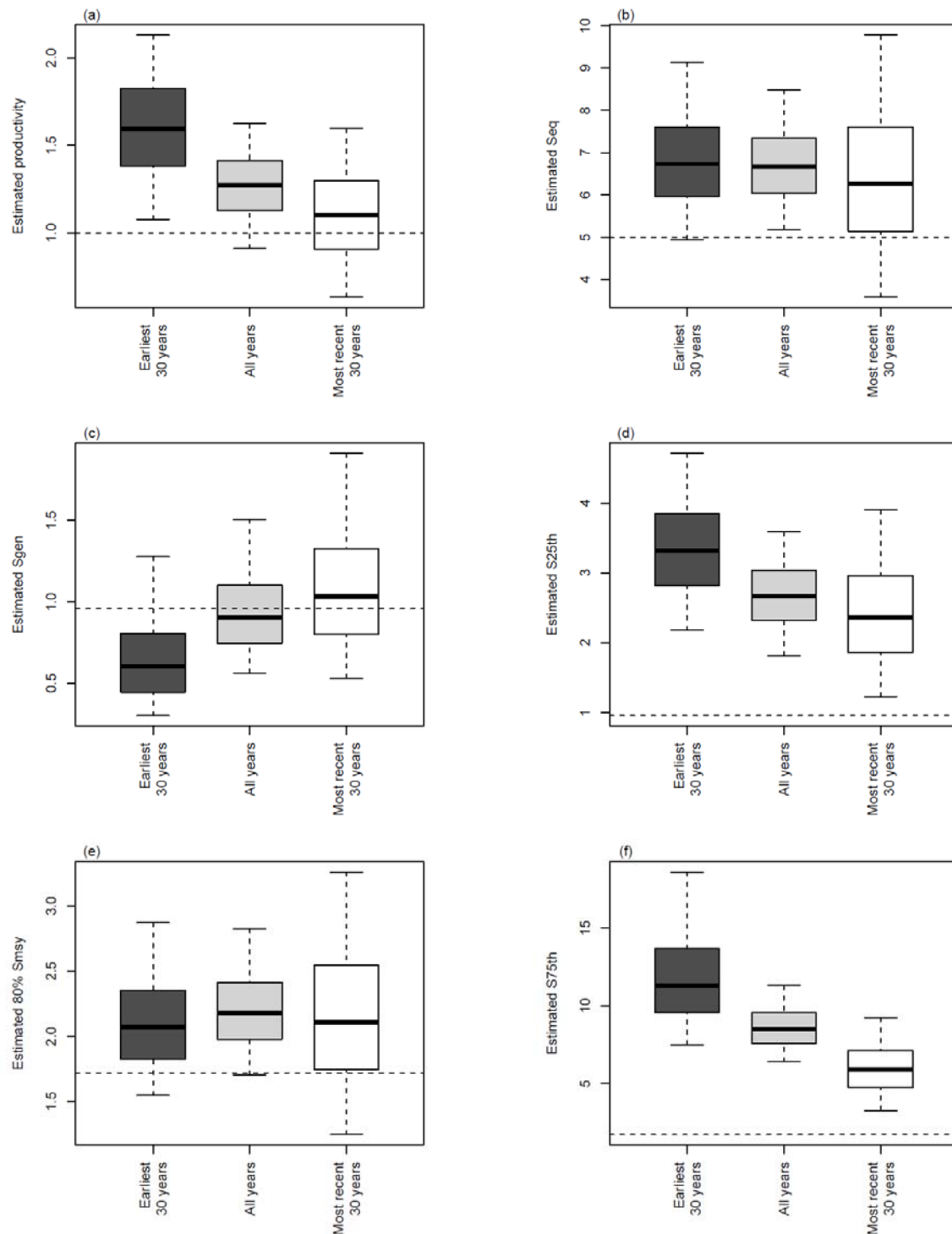


Fig. 15. Mean percent error of estimated Ricker parameters and lower and upper benchmarks from the “true” values, using only the first 30 years of data (black bars), all 70 years (grey bars), and the most recent 30 years (white bars) under the assumption that  $S_{eq}$  remained constant over time (a), or  $S_{max}$  remained constant (b).

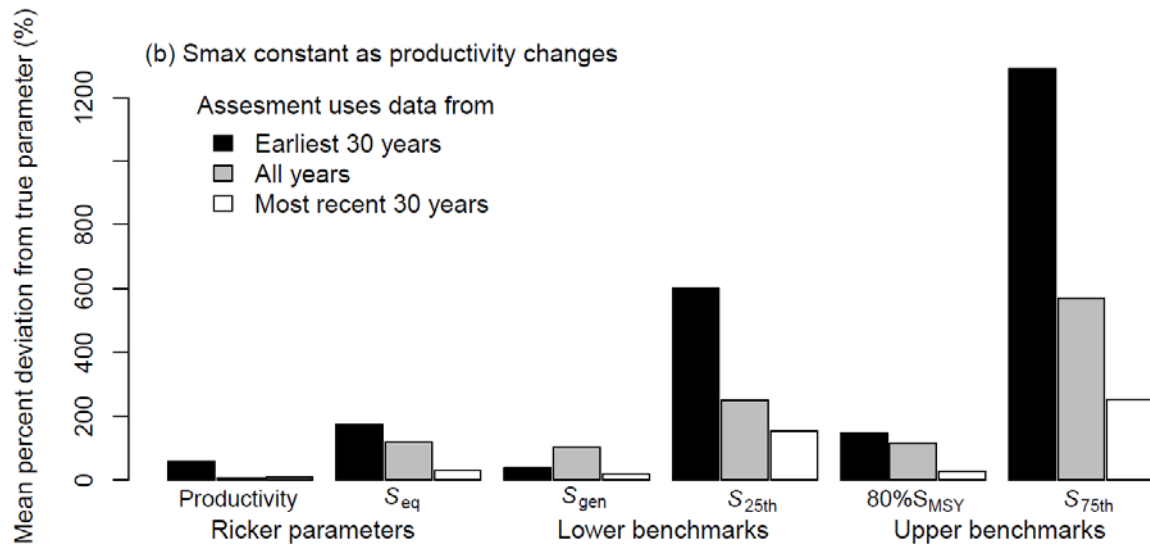
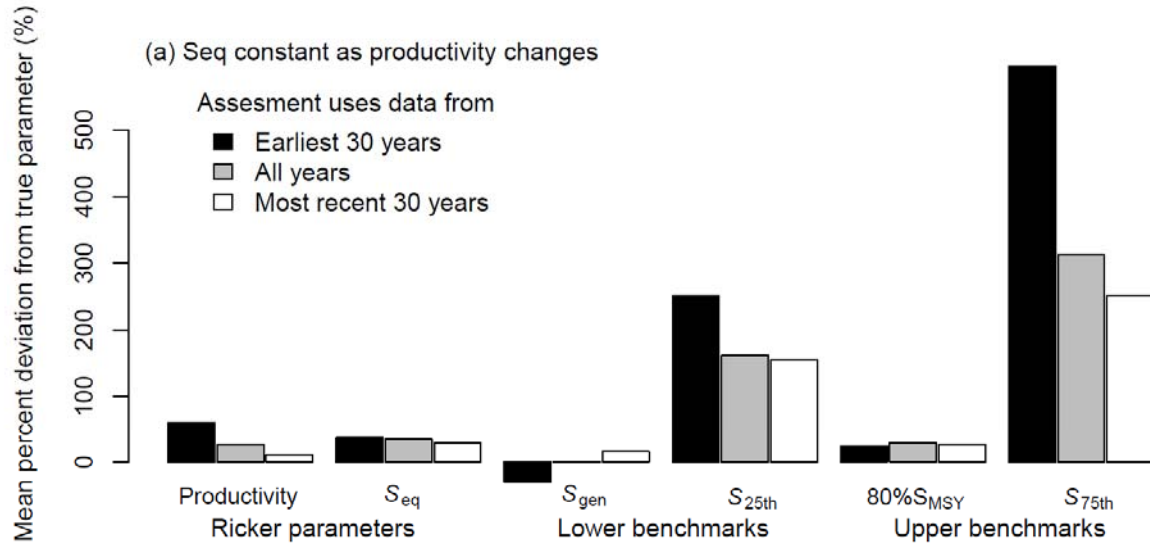
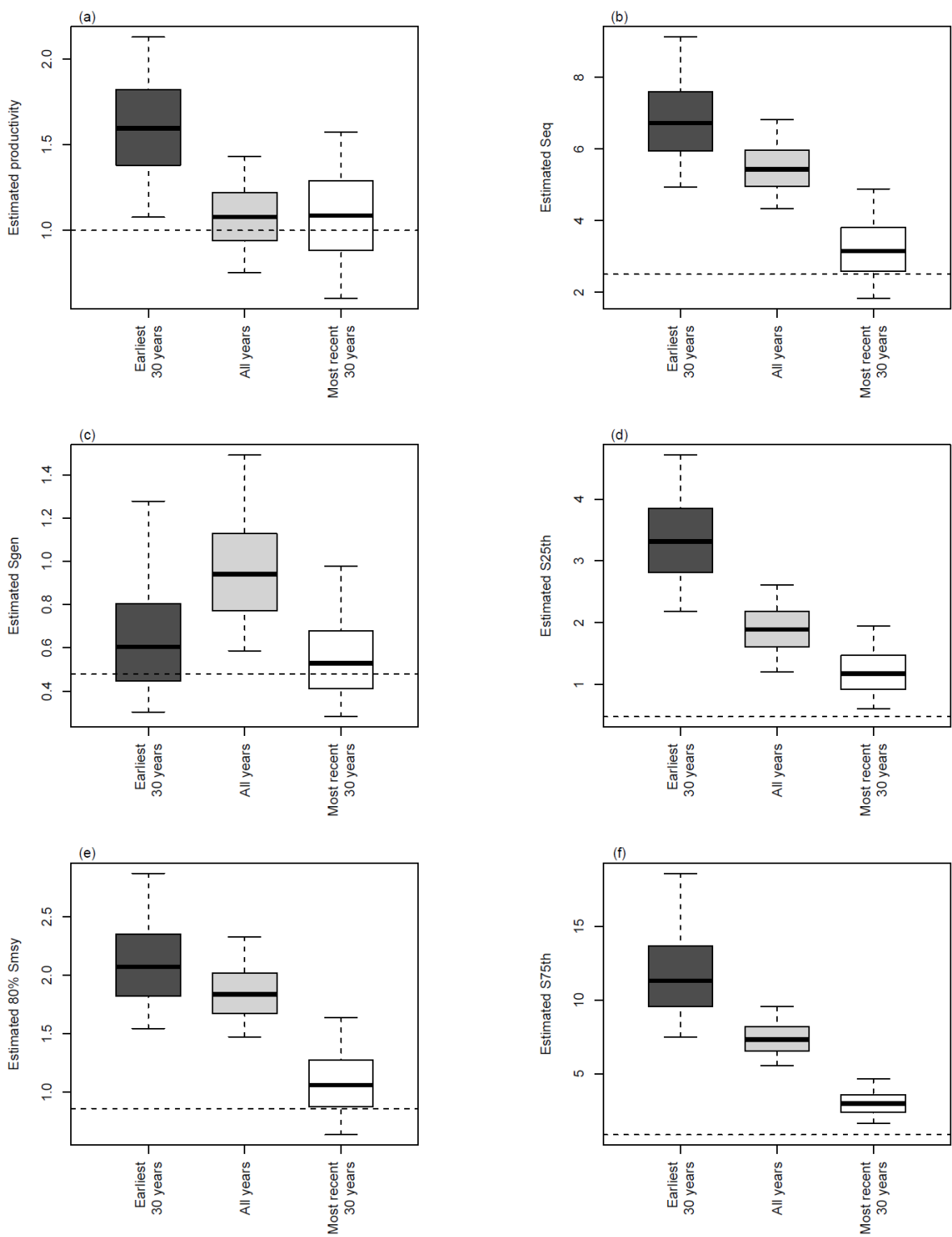


Fig. 16. Caption as for Fig. 14, except  $S_{\max}$  was held constant and  $S_{\text{eq}}$  declined in simulations as productivity varied.



## Tables

Table 1. Parameter and benchmark estimates and upper/lower credible interval bounds delineated as 2.5th and 97.5th posterior densities for most recent year, 2012.

	South Coast Streams				Northeast Vancouver Island				Upper Knight			
Model	Standard		Hierarchical		Standard		Hierarchical		Standard		Hierarchical	
Statistic	Estimate	UCL/LCL	Estimate	UCL/LCL	Estimate	UCL/LCL	Estimate	UCL/LCL	Estimate	UCL/LCL	Estimate	UCL/LCL
Ricker $\alpha$	1.39	2.23	1.59	2.35	1.54	2.28	1.66	2.51	2.20	4.06	2.14	3.53
		0.92		1.03		1.06		1.24		1.22		1.30
$S_{\max}$	80,350	225,841	68,054	191,642	117,170	317,526	104,053	190,717	16,675	61,095	16,950	61,125
		43,576		41,660		67,393		60,788		9,468		9,885
$S_{\text{gen}}$	9,665	20,254	10,018	18,720	16,589	32,035	16,386	25,983	2,991	8,944	3,086	9,711
		1,925		3,939		7,668		10,848		1,521		1,712
	10,400	20,580	11,553	20,915	18,631	32,969	19,455	30,422	4,600	11,844	4,572	12,322
		1,572		3,302		6,458		10,879		2,133		2,273
	Loughborough				Bute Inlet				Georgia Strait			
Model	Standard		Hierarchical		Standard		Hierarchical		Standard		Hierarchical	
Statistic	Estimate	UCL/LCL	Estimate	UCL/LCL	Estimate	UCL/LCL	Estimate	UCL/LCL	Estimate	UCL/LCL	Estimate	UCL/LCL
Ricker $\alpha$	2.27	3.26	2.24	3.10	2.47	3.67	3.45	3.55	3.07	4.80	2.71	4.29
		1.55		1.61		1.59		1.74		2.06		1.96
$S_{\max}$	62,852	124,314	64,706	121,123	107,752	266,610	267,472	194,547	496,295	1,074,352	593,563	1,128,389
		43,882		44,356		70,210		69,239		297,467		327,131
$S_{\text{gen}}$	12,002	21,221	12,316	21,481	20,528	47,014	48,072	35,976	91,724	203,578	113,305	214,955
		8,118		8,308		12,420		12,474		44,305		52,284
	18,219	26,700	18,301	28,150	33,752	61,523	64,063	52,345	187,546	302,108	201,020	316,844
		13,411		13,599		23,483		23,013		141,771		147,089
	Howe Sound to Burrard Inlet											
Model	Standard		Hierarchical									
Statistic	Estimate	UCL/LCL	Estimate	UCL/LCL								
Ricker $\alpha$	2.63	3.75	2.47	3.49								
		1.81		1.79								
$S_{\max}$	511,173	1,657,672	559,155	1,837,905								
		308,310		333,798								
$S_{\text{gen}}$	97,554	310,845	107,571	344,097								
		54,892		60,229								
	171,126	410,187	177,421	453,338								
		119,131		120,094								

Table 2. Benchmark values across three methods used: standard Ricker model, hierarchical Ricker model ( $S_{gen}$  and 80%  $S_{MSY}$ ) and percentiles (25<sup>th</sup> and 75<sup>th</sup>).

		Southern Coastal Streams		
	Method	Standard Ricker	Hierarchical Ricker	Percentile
Upper Benchmark		10,400	11,553	54,350
Lower Benchmark		9,665	10,018	5,425
		North East Vancouver Island		
Upper Benchmark		18,631	19,455	75,136
Lower Benchmark		16,589	16,386	16,519
		Upper Knight		
Upper Benchmark		4,600	4,572	11,191
Lower Benchmark		2,991	3,086	2,006
		Loughborough		
Upper Benchmark		18,219	18,301	46,303
Lower Benchmark		12,002	12,316	17,313
		Bute Inlet		
Upper Benchmark		33,752	33,247	85,517
Lower Benchmark		20,528	21,155	11,275
		Georgia Strait		
Upper Benchmark		187,546	201,020	445,139
Lower Benchmark		91,724	113,305	202,269
		Howe Sound to Burrard Inlet		
Upper Benchmark		171,126	177,421	303,280
Lower Benchmark		97,554	107,571	85,394



Table 3. Conservation status for each CU for the most recent year of analysis, 2012. Statuses are calculated using all data available up to return year 2012 (brood years 1955-2006) for the Ricker-based benchmarks, and use all escapement data (1953-2012) for the percentile-based benchmark.

Conservation Unit	Percentile Status	Standard Ricker Status	Hierarchical Ricker Status
South Coast Streams	Red	Red	Red
Northeast Vancouver Island	Amber	Green	Green
Upper Knight	Amber	Green	Green
Loughborough	Amber	Green	Green
Bute Inlet	Amber	Green	Green
Georgia Strait	Green	Green	Green
Howe Sound to Burrard Inlet	Green	Green	Green

Table 4. Proportion of years where Ricker-based status and percentile-based status match, by CU and Ricker Model (standard Ricker model in column 1 and hierarchical Ricker model in column 2). Columns 3 and 4 show the proportion of years where the percentile-based status matched OR was more precautionary than Ricker-based status.

Conservation unit	Standard Ricker match with percentile benchmarks	Hierarchical Ricker match with percentile benchmarks	Standard Ricker match or more precautionary	Hierarchical Ricker match or more precautionary
South Coast Streams	0.27	0.41	1.00	1.00
Northeast Vancouver Island	0.19	0.19	0.96	0.97
Upper Knight	0.49	0.51	0.77	0.77
Loughborough	0.21	0.21	1.00	1.00
Bute Inlet	0.58	0.58	0.77	0.77
Georgia Strait	0.28	0.28	1.00	1.00
Howe Sound to Burrard Inlet	0.56	0.56	1.00	1.00

Table 5. Parameters used as base case, univariate sensitivity analyses, and global sensitivity analyses of simulation model to evaluate lower benchmarks.

Sub-model	Parameter	Base-case Value	Values considered in univariate sensitivity analyses	Range considered in global sensitivity analyses
Population dynamics sub-model	Ricker productivity parameter	1	0.5 (low) and 2.0 (high)	0.5-2.0
	Ricker autocorrelation coefficient	0.6	0 (low) and 0.9 (high)	0-0.9
	Standard deviation in Ricker residuals	0.75	0.6 (low) and 1.0 (high)	0.6-1.0
	Average proportions at age-of-maturity	Age 3=24%; Age 4=67% Age 5=9%		
	Natural variability in age-at-maturity, $\sigma_n$ , specified in a multivariate logistic distribution	0.8	0.1 (low) and 0.9 (high)	0.1-0.9
	Correlation in recruitment residuals among subpopulations within a CU	0.4	0 (low) and 1.0 (high)	0-1.0
	Initial spawner abundances	$0.2 \times S_{eq}$ , where $S_{eq}$ is spawner abundances at equilibrium	$0.1 \times S_{eq}$ (low) and $0.3 \times S_{eq}$ (high)	$0.1 \times S_{eq}$ - $0.3 \times S_{eq}$
	Stray rate	0.05		
Observation sub-model	Variability in observed age-at-maturity, $\sigma_n$ , specified in a multivariate logistic distribution	0.1	0 (low) and 0.9 (high)	0.1-0.9
	Standard deviation in observation errors of spawner abundances	0.5	0.2 (low)	0-1.0
	Standard deviation in observation errors of catches	0.5	0.2 (low)	0-1.0
	Multiplicative bias in observed spawner abundances not accounted for in assessment	1	0.8 (negative bias) and 1.2 (positive bias)	0.8-1.2
Assessment sub-model	Proportion of subpopulations sampled within a CU	100%	50% (low)	50%-100%
	Proportion of years that CU is sampled	100%	60% (low)	60%-100%

Harvest sub-model	Harvest rate during initialization period	20%	10% (low) and 50% (high)	10%-60%
	Outcome uncertainty (standard deviation in differences between target and realized harvest rates)	0.3	0.5 (high)	0-0.9

## Appendix

### Objective 1a: Parameterization of priors on $S_{max}$ for the standard Ricker model

For the parameterization of the uniform prior we assumed that  $S_{max}$  was less than twice the maximum observed spawner value, which is likely given the observed low to moderate harvest rates on average for most Inner South Coast chum salmon CUs, with the possible exception of Georgia Strait. For parameterization of the log-normal prior, we set the width of lognormal prior by using a CV of 5, which we found to produce priors in which the highest probability values occurred in approximately the same range of  $S_{max}$  as the uniform distribution. The lognormal prior is weakly informative, as it pulls posterior distributions of  $S_{max}$  towards mean observed escapement. Although most of the weight of the prior distribution lies within the same range as the uniform distribution, it also includes values of  $S_{max}$  far greater than the observed spawner levels. Therefore, using a log-normal prior distribution, some posterior estimates of  $S_{max}$  may be far higher than the range of historically observed escapement, which may be the case if the CU had been long suppressed far below historical levels.

**Objective 1b: Evaluation benchmarks using simulation analyses- results assuming  $S_{\max}$  constant given variability in productivity**

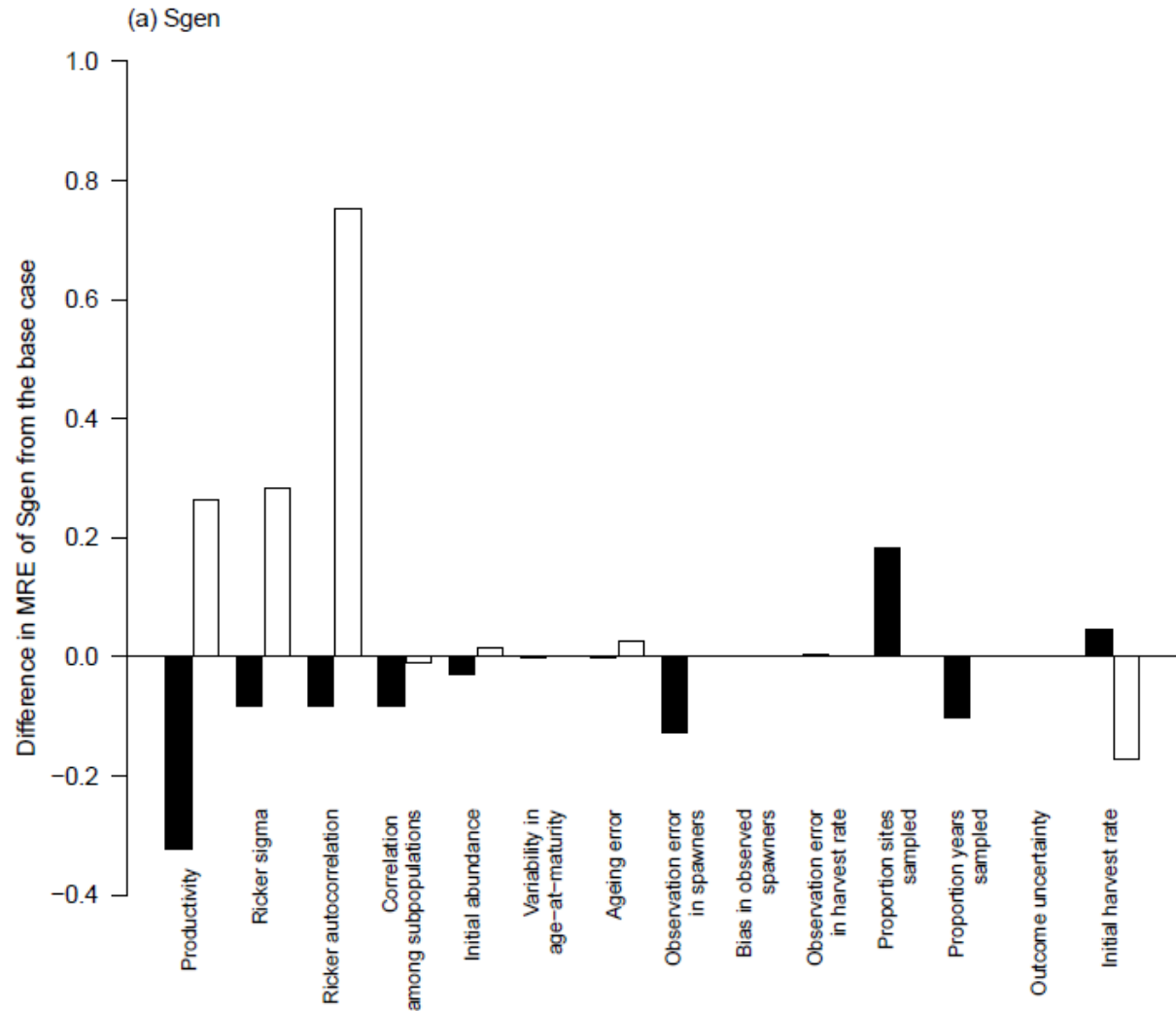


Fig. A1. Difference in the mean raw error, MRE, of estimated lower benchmark ( $S_{\text{gen}}$ , (a), and  $S_{25\text{th}}$  (b)) and the “true” lower benchmarks (“true”  $S_{\text{gen}}$ ), between sensitivity analyses listed on the X-axis and the base case scenario. Black bars are analyses where the input parameter was increased relative to the base case (see Table 1); white bars are analyses where the input parameter was reduced relative to the base case. Positive values indicate sensitivity analyses where MRE increased under that change in input parameter from the base case; negative values indicate analyses where the MRE declined under that change in input parameter.

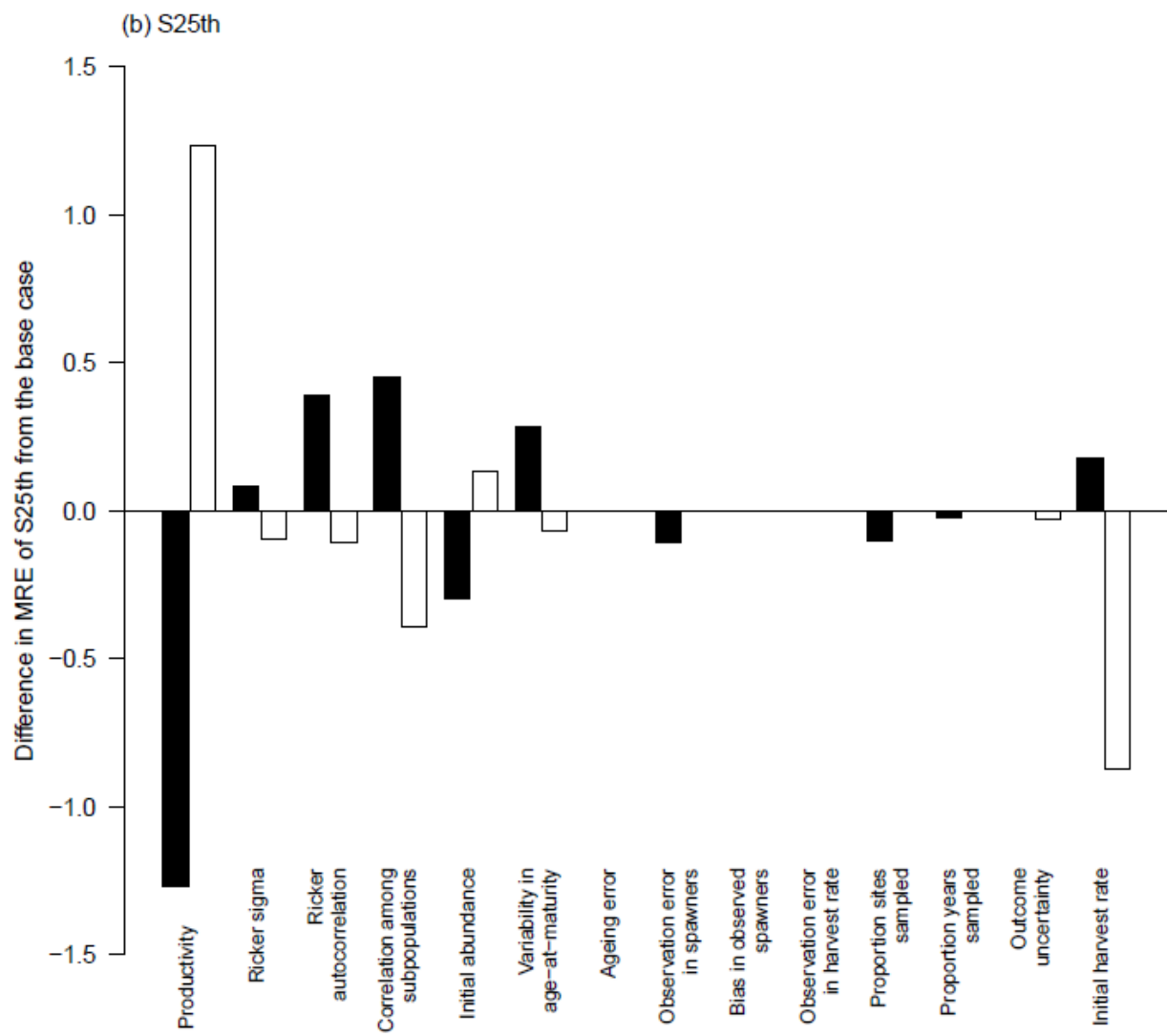


Fig. A1 cont.