

# **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>2</sup>, Diana Dobson<sup>3</sup>, Joe Tadey<sup>2</sup>, Pieter Van Will<sup>4</sup>, Lyse Godbout<sup>1</sup>, and Wilf Luedke<sup>3</sup>

<sup>1</sup>Pacific Biological Station, DFO, Nanaimo, BC; <sup>2</sup>DFO, Annacis Island, New Westminster, BC; <sup>3</sup>DFO, South Coast Area Office, Nanaimo, BC; <sup>4</sup>DFO, Port Hardy, 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-recruitment models to calculate benchmarks and assign green, amber, or red status to CUs. Data limitations for many CUs require the exploration of alternative benchmarks to ensure conservation objectives are achieved when stock-recruitment data are not available. In this study we compare the performance of alternative lower and upper benchmarks for data-limited CUs based on 25<sup>th</sup> and 75<sup>th</sup> percentiles of observed abundances, 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 were generally more precautionary than previously adopted stock-recruitment based benchmarks for the 9 CUs and 5 stock management units of Chum Salmon analyzed here. The simulation study yielded similar results. However, when population productivity was moderate to low and harvest rates were high, percentile-based lower benchmarks tended to be below "true" lower benchmarks. In those cases, we recommend a higher percentile be applied as lower benchmark, such as 50<sup>th</sup> percentile, instead of the 25<sup>th</sup> percentile. We further provide recommendations on when percentile benchmarks should not be applied, based on estimates of productivity and harvest rates.

## **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.

In the first year of this two-year project, we evaluated benchmarks for data-limited, percentile-based benchmarks against data-rich benchmarks for CUs of Chum Salmon on the Inner South Coast of BC. Specifically, we identified data-limited percentilebased benchmarks and data-rich Ricker-based benchmarks (which require stock-recruitment data) for those CUs, and evaluated their performance in a retrospective analysis. We further evaluated performance in a prospective simulation model under various hypotheses about productivity (among other sources of uncertainty). In that analysis, we found that percentile-based benchmarks tended to be more precautionary than data-rich benchmarks, except at low productivity and high initial harvest rates. However, that analysis was limited to one region (excluding West Coast of Vancouver

Island and Fraser River), and had two technical limitations in the simulation model related to back-transformation bias associated with log-normal error distributions and application of harvest rates. In this report we describe updated modelling results with the new data. We provide context and model description as in Year 1's report for completeness.

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 research objectives were to:

- (1) Develop hierarchical models of stock-recruitment data that combine information across numerous CUs within the west coast of Vancouver Island and Fraser River areas. This approach can be robust to uncertainties in underlying data than standard single-CU stock-recruitment analyses.
- (2) Compare performance of percentile-based benchmarks with single-CU and multi-CU (hierarchical) stock-recruitment based benchmarks using (a) retrospective analyses of empirical data and (b) simulation modelling that accounts for the high uncertainties chum spawner, catch, and recruitment abundance estimates.
- (3) Provide recommendations on the application of benchmarks to chum management units (through component CUs within management units) and Chum Salmon Genetic Units, GUs (as identified by a project funded by the PSC SEF on genetic stock identification, J. Candy) on the west coast of Vancouver Island on in the Fraser River within the context of the PST Chum Annex.

To implement objective 1, we considered deriving recruitment estimates for Fraser River and Inner South Coast from a new run reconstruction model ChumGEM developed by SFU and the PSC Chum Technical Committee. The goal for that model was to generate return size estimates that are more rigorous than those currently available from a spread-sheet based backwards run reconstruction. We reviewed this model, revised several components, and explored sensitivity of run sizes to various input parameters (details below). Sensitivities of run size estimates to input parameters and several questionable assumptions precluded us from using this model further.

Genetic Units are informed by microsatellite DNA data from limited sampling of Chum Salmon in BC, focusing primarily on Strait of Georgia, and tend to be coarser in resolution than CUs for Chum Salmon in southern BC. With limited sampling in more northern areas, differentiation of GUs on the west coast of Vancouver Island was not possible, nor was resolution north of the Strait of Georgia for the Inner South Coast. Although Genetic units are the basis for run reconstruction model, ChumGEM, this model was not used for stock-recruitment modelling here, and so we have focused our results on CUs, the unit of biological diversity recommended under Canada's Wild Salmon Policy.

After reviewing data availability for Fraser River further, those CUs were removed from the retrospective analyses due to insufficient information on spawner counts and recruitment. We

focused instead on the west coast of Vancouver Island CUs (Southwest Vancouver Island and Northwest Vancouver Island CUs) and 5 component stock management units within the Southwest Vancouver Island CU. The stock management units we considered within Southwest Vancouver Island CU were: Barkley, Clayoquot, Nootka, Esperanza, and Kyuquot. The simulation modelling results, however, are generic for Chum Salmon CUs, including Fraser River.

We first describe the application of hierarchical stock-recruitment models to both Inner South Coast (from year 1 of this study) and West Coast of Vancouver Island (year 2), in the context of retrospective analysis evaluating data-rich and data-limited benchmarks. In this way, we have combined the Methods and Results for Objectives 1 and 2a in this report. We follow with a description Methods and Results for Objectives 3, simulation evaluation of benchmarks, and conclude with recommendations on application of these benchmarks.

### **Context: 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" (DFO 2005).

For populations with time-series of stock-recruitment 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-recruitment 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-recruitment 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. (2014) 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. (2014) 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 ADFG (Clark et al. 2014).

In this report, we have combined Methods and Results for Objective 1 and 2a, the development of standard Ricker and hierarchical stock-recruitment models for Westcoast Vancouver Island CU and component stock management units, and the retrospective evaluation of data-limited benchmarks against data-rich versions that use stock-recruitment data (both standard Ricker and hierarchical models). This is followed by Methods and Results for Objective 2b, an evaluation of data-limited against data-rich benchmarks in simulation. We further describe updated analyses of benchmarks that use truncated data under scenarios of time-varying productivity (i.e., only recent data to reflect current conditions or only historical data to avoid shifting baseline). We conclude with recommendations of applications of data-limited benchmarks under gradients in productivity and harvest rates (Objective 3), and a discussion of impacts of time-varying productivity.

## **Objective 1. Development of hierarchical stock-recruitment models for West Coast of Vancouver Island and Objective 2a. Evaluating benchmarks using retrospective analyses**

The goal of our retrospective analysis was to compare status reached under both data-rich and data-limited biological benchmarks. 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 or CU components are assumed to have productivity values which are “drawn” from a shared distribution, centered on an overall mean productivity rate. Hierarchical models may reduce uncertainties and biases in parameter estimation mentioned above by sharing information on productivity across populations, 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 using a recursive Bayes modelling approach. Although results for Inner South Coast Chum Salmon CUs were presented in the Final Report for Year 1 of this project, they are provided again here to compare against new results for the West Coast of Vancouver Island.

### **Data**

#### ***Inner South Coast (ISC) Chum Salmon data***

Historical time-series of escapement and returns were available for seven CUs of Inner South Coast Chum Salmon. Escapement data, identified as either wild or hatchery, were available for these CUs from 1953-2012, while CU-specific return data were reconstructed from exploitation rates, migration timing and patterns, spawner abundances, and age distributions, for brood years 1955-2006 (P. van Will pers. comm. 2016). Time series’ of returns were generated from a run reconstruction model that used 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 (Van Will 2014). Wild recruitment was estimated by assuming that proportions of wild fish in catches were equal to the proportion in observed escapement. Historical genetic composition of the catches was not available to identify the wild and hatchery contribution to the catches.

In years where spawner abundances (escapement) were missing, data have been infilled using standard approaches assuming covariation in abundance trends across sites within CUs (Van Will 2014). On average, across CUs and years, 45% of sampling sites were surveyed (ranging from 27% for the Howe Sound–Burrard Inlet CU to 57% for the Bute Inlet CU). 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), assuming covariation in abundance trends among CUs. Infilled escapement and return data were combined with age-composition data to create brood tables – from which a stock-recruitment time series was formulated.

Fitting the Ricker model to uncertain data can lead to biased parameter estimates because of observation errors in escapement (i.e., errors-in-variables) and time-series biases (Walters and Martell 2004). These time-series are relatively long (51 years) and contrast in escapement observations is high (ratio of maximum to minimum spawner abundances ranged from 8-2600, mean=481), which should ameliorate these biases (Walters and Martell 2004). However, caution

in the interpretation of results is warranted, and these results should be considered in conjunction with those from simulation model that incorporates multiple sources of data uncertainties (See Objective 2b for a more thorough description of time-series biases).

### ***West Coast of Vancouver Island, WCVI, Chum Salmon***

Time series of escapement and returns from 1953-2015 were available for both the Southern and Northern West Coast Vancouver Island CUs (SWVI, NWVI), with the SWVI CU being split into five stock management units (SMUs). Brood tables were constructed based on yearly age composition data at the SMU level. When gaps were found in aging data, they were infilled based on surrounding SMU's with data for that year. Aging data begins in 1959, with a gap in the mid-1960's resulting in stock-recruitment data from 1956-2010, with a 6-year gap in the 1960's. There has been hatchery production in this area since the 1970's, and the data used in this analysis excludes these hatchery populations. Nitinat (Area 22) and Tlupana (Area 25) were removed, as they are dominated by hatchery populations. The proportion that Tlupana contributed to catches in Areas 25 was estimated from marking data and terminal fisheries. Nitinat is the only population within Area 22. Similar to the ISC data set described above, the WCVI CUs and SMUs have relatively long time-series (~50 years, after accounting for gaps) and show considerable contrast (max/min spawner ratio ranged from 9-26, mean=18), which may ameliorate biases associated with errors in spawner abundances and time-series biases. However, due to infilling and assumptions made in run reconstruction, caution should be taken in the interpretation of results.

### ***Fraser River***

Current recruitment estimates for Fraser River Chum Salmon (both Fraser, and Fraser Canyon CUs) are not available. Although historical stock-recruitment time-series have been published for the Fraser River (Ryall et al. 1999, brood years 1959-1994), experts familiar with the system deem those data unreliable (J. Tadey, pers. comm. 2016), despite their use in at least one peer-reviewed study (Malick and Cox, 2016). Additionally, the model used to generate the recruitment time-series has not been maintained in recent years.

### ***Chum Genetic and Environmental Management model, ChumGEM***

We explored a forward run reconstruction model developed for Chum Salmon in southern BC and Washington (the Chum Genetic and Environmental Management model, ChumGEM) as source of recruitment time-series for Fraser River CUs and to validate recruitment time series for ISC CUs. ChumGEM was developed by Simon Fraser University under the direction of the Pacific Salmon Commission Chum Technical Committee in 2015. ChumGEM was built using R and ADMB code, imbedded within a graphic user interface (GUI). The application also serves as a data repository for escapement, catch (commercial/test), and genetic stock identification (GSI) data. Despite being delivered to DFO in 2005, the model has not yet been validated, limiting its value in providing recruitment estimates. Here we briefly describe the mode, and our evaluation of the sensitivity of model outputs (recruitment estimates) to uncertainties in input parameters, as a way of validating the model.

The ChumGEM model uses data on catch (commercial and test fisheries), recent genetic stock identification of catch, test fishery CPUE, and escapement by stock to estimate return size by stock. Due to constraints of genetic stock identification (GSI) data, populations are modeled at

the Genetic Unit scale (GU) instead of at the CU scale. GUs for ISC, WCVI, Fraser River, as well as Washington state, USA, are modelled during their return migration from the northern tip of Vancouver Island, through fisheries on either the west or east coast of Vancouver Island, to their entry into freshwater by sequentially moving fish southward through numerous spatial fishing areas (Fig. 1). Arrival timing to the northern tip of Vancouver Island is modelled with a normal distribution, with informative priors on associated parameters. Fish are then stepped through fishing areas with one-day increments. Fish move according to pre-determined probabilities of migrating along either east coast of Vancouver Island (diversion rate), and pre-determined swim speeds. Probabilities of moving from one fishing area to another in a given day are estimated using an ordered multinomial logit model based on swim speeds. These probabilities are calculated prior to model fitting. Catches are removed from return abundances in proportions informed by observed GSI data.

In our preliminary review of the model, we identified at least four assumptions that are likely violated or unrealistic. We briefly describe our review here; more details are available in Davis (2016).

- (1) In the model, after fish are directed around either side of Vancouver Island according to given diversion rates, abundances are divided equally among remaining migration routes from each fishing area. However, the proportion of fish assigned to each migration route likely varies as some routes are more well-used than others (i.e., they are not equally used). The probability associated with each migration route could instead be informed by expert opinion and/or available data.
- (2) Additionally, we found that model estimates of GU-specific recruitments were extremely sensitive to assumptions about the diversion rate. Currently, the diversion rate is assumed to be constant over GUs and time, and does not reflect our current understanding of variability in that rate.
- (3) The model applies informative priors on recruitment. We found that the posterior estimates of recruitment (model outputs) were sensitive to the form of the priors, at least for some GUs. Further efforts to identify plausible GU-specific priors and/or a model reformulation are required to reduce model sensitivity to prior assumptions.
- (4) The model, in its current form is also limited to years with existing GSI data (2008-present). The model was designed as a tool for post-season run reconstruction, and therefore fits years individually. An extended, multi-year version of this model could be developed to generate historical time series of recruitment. A multi-year model may also improve parameter estimates by borrowing information across years. For example, migration timing is likely similar across years, and a multi-year model could use migration timing estimates from years with high quality data to inform years with low quality data. Additionally, if estimates of average migration timing across years could be estimated, it may be possible to use catch and escapement data to estimate returns for years without GSI data to reconstruct recruitment time series.

Due to unrealistic assumptions and priors that require more scrutiny, we were unable to develop reliable recruitment estimates from ChumGEM. Based on our analyses and recommendations presented to the PSC Chum Technical Committee (Davis 2017), the committee has expressed their commitment to continue revising and validating the run reconstruction model. However,

reconstructed run sizes for Fraser River will not be available to use as a case study for this project.

Therefore, we focused our analyses, on the 7 Inner South Coast CU's and two West Coast Vancouver Island CUs described above. As an additional case study on finer scale data, we implemented the same retrospective analysis on 5 component SMUs within the Southwest Vancouver Island CU.

## Methods

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 both stock-recruitment based and 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. 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 5 or 6 years (for ISC and WCVI, respectively), depending on the stock. Therefore, stock-recruitment based benchmarks and statuses were calculated for years 1970-2012 for ISC, and 1976/1977-2015 for WCVI. These benchmarks were calculated using parameters from Ricker models fit using data from brood years 1964-2006 for ISC, and 1957-1958, 1965/1966-2010, for WCVI. Lower and upper benchmarks were compared to generational mean escapements to determine status. Generational mean escapement was estimated as the four-year running geometric average.

### *Standard Ricker Model*

The standard Ricker and Hierarchical models were developed in Year 1 of the project, and are described below. For each year with sufficient data, a standard 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  $\alpha$  (also referred to as 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 (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) (See Appendix A for plots of priors and posteriors of  $\alpha$  parameter),

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

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 spawner 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 B 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 CU's within the two groupings (ISC and WCVI) 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  $\sigma_\alpha = 1/\tau_\alpha$ ,

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

The Southwest Vancouver Island CU, the hierarchical model was also run for SMUs within that CU, assuming productivity parameters for each SMU were drawn from a common distribution. Models were fit using MCMC runs 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),

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

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 ( $S_{25th}, S_{75th}$ ). Holt and Ogden (2013) recommended against using stock-recruitment benchmarks when Ricker  $\alpha$  falls below 1.5. We have removed years when  $\alpha < 1.5$  from our retrospective analysis.

#### *Retrospective Analysis*

Retrospective analysis was carried out by stepping through each year with sufficient data (at least 10-year time-series) and then estimating benchmarks and assessing status using all available data up until that year. This mimics the analysis that would have been carried out, and the status reached if these benchmarks had been used in the past. Since percentile benchmarks are defined as the 25<sup>th</sup> and 75<sup>th</sup> percentiles of historical spawner abundances (which are provided as “data”), these values do not have associated uncertainties. However, Ricker-based benchmarks are calculated based on model parameters, and have associated uncertainties. We used two slightly different approaches to characterize the uncertainty in benchmarks, and the resulting uncertainty in status assessments. In order to properly assess uncertainty, and to account for the widely documented negative correlation between Ricker parameters, we estimate Ricker benchmarks for each MCMC “draw”. This allows the estimation of benchmarks based on pairs of Ricker parameters from each MCMC draw, rather than the median and bounds of each, calculated over all MCMC draws. From these draws we can express each benchmark as a median, with 95% credible intervals, estimated as the 2.5% and 97% posterior densities of each benchmark. In the second approach, we estimated benchmarks for each sample from the posterior distribution of parameters, and a corresponding generates a status assessment against those benchmarks. This means that for each year, we generated a probability estimate associated with each status (red, amber, green). In other words, we estimated the probability that the CU has each of the three statuses for each year.

#### *Changes in productivity*

To identify changes in productivity over time for Chum Salmon CUs 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, but 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,

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

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*

For all CUs and SMUs, 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. However, these differences were small and estimates consistently fell within the range of uncertainty under the alternate assumption (Figs. 2 and 3 show the most recent parameter estimates and CUs and SMUs, respectively). Furthermore, when comparing statuses, models fit with either prior matched between 85-100% of years depending on CU/SMU, and therefore do not appear to make a significant difference in the assignment of status. We report results using uninformative uniform priors here.

### *Current benchmarks and status*

Lower percentile benchmarks ( $S_{25th}$ ) tended to be similar or higher 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%  $S_{MSY}$ ) (Figs. 4 and 5, for CUs and SMUs respectively, Table 1). The Ricker-based benchmark,  $S_{gen}$ , has the characteristic of being relatively high when productivity is low (i.e., is precautionary when conditions are poor) and being low when productivity is high (Holt and Folkes 2015). Our results support this finding. In particular, we found that the  $S_{25th}$  benchmark tended to be much higher than  $S_{gen}$  when productivity was high, and this difference was reduced when productivity was low (Appendix C).

Stock-recruitment benchmarks varied slightly between the standard and hierarchical Ricker models (comparing Fig. 4, top and bottom panels for each CU), but these differences were small compared with the large uncertainties in benchmark estimates (Table 2). The posterior distributions of the upper and lower benchmarks,  $S_{gen}$  and 80%  $S_{MSY}$ , overlapped, and in some cases were nearly indistinguishable, e.g., Southern Coastal Streams and North East Vancouver Island (Fig. 4a and b).

In retrospective analyses of the hierarchical model compared with the standard Ricker model, we found that uncertainties in estimates of  $\alpha$  and  $S_{max}$  (Ricker parameters) were reduced slightly for the hierarchical model in some CUs on the Inner South Coast. The hierarchical Ricker model

tended to reduce uncertainties for those CUs and years where productivity estimates were similar across CUs (Fig. 2). Alternatively, when productivity estimates from a CU differed from neighbouring CUs, uncertainty bounds tended to increase (e.g., Southern Coastal Streams and Northeast Vancouver Island CUs, Fig. 2). For the WCVI CUs, uncertainty around productivity estimates did not differ between hierarchical and basic Ricker models (Fig. 2).

Statuses for the most recent year for which status could be assessed (2012 for ISC CUs, 2015 for WCVI CUs and SMUs) determined using all data available up to that return year (and using brood years up to 2006/2010) are shown in Table 3. Percentile-based statuses were the same or more precautionary than Ricker-based statuses in that 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 for ISC CUs due to high contrast in time-series (Fig. 6, bottom row compared to first two rows of panels). For WCVI CUs and SMUs, which did not exhibit large contrast, percentile benchmarks were more consistent over time (Fig. 6 for WCVI CUs and Fig. 7 for SMUs).

Stock-recruitment benchmarks tended to remain relatively consistent over time for four CUs (Upper Knight, Loughborough, SWVI, NWVI), 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. 6). Uncertainties in stock-recruitment benchmarks tended to decline over time for 3 ISC CUs (Southern Coastal Streams, Upper Knight, and Loughborough) and all WCVI CUs and SMUs, but 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 (Southern Coastal Streams, North East Vancouver Island, and Upper Knight); the others remained constant (Loughborough, SWVI, NWVI, all WCVI SMUs) 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.

Large uncertainties in stock-recruitment benchmarks resulted in uncertainties in status assessments, which we present in two ways, as described above. First, benchmarks are represented as the median and 95% posterior densities, representing the expected value and the 95% credible interval (Fig 8, 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.

8a). The probability of a CU having each status, in each year can be seen in Figures 10 and 11 (colours on vertical bars depicting probability of each status). Each probability is associated with the proportion of MCMC draws which yielded that status, in that given year.

### ***Comparing Benchmarks***

Percentile benchmarks were found to provide the same, or more precautionary status compared to Ricker-based benchmarks (Figs 8 and 9, 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 for ISC CUS (Table 4). Alternatively, for WCVI CUs and SMUs, percentile and Ricker-based benchmarks rarely matched (0-10%). On average, the percentile benchmark provided the same or more precautionary status in 94% of years for both model types across CUs (Table 4). For SWVI SMUs, status assessed from percentile benchmarks were the same or more precautionary 100% of the time. 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 resulting in declining  $S_{25th}$  benchmarks, paired with relatively constant  $S_{gen}$  values (e.g., Upper Knight from 1999-2001, see Fig. 4c), 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. 4e).

The two Ricker-based benchmarks (standard and hierarchical Ricker) gave the same status 98% of years when averaging across CUs, and 100% of years for SWVI SMUs.

### ***Productivity over Time***

Temporal patterns in productivity varied across CUs and SMUs (Fig. 12). Declines in productivity over time were observed in three CUs (Southern Coastal Streams and Loughborough, SWVI, Fig. 12a,d, and h), increases followed by declines in three CUs (North East Vancouver Island, Bute Inlet, and Georgia Strait, Fig. 12b, e, and f) and consistent levels followed by a small increase in Howe Sound to Burrard Inlet and NWVI (Fig. 12g and i). Estimates of productivity for Upper Knight (Fig. 12c), were highly variable and uncertain. WCVI SMUs all show slight declines over time, though declines are small compared with uncertainty in annual productivity estimates (Fig. 13). There was considerable uncertainty in productivity for all CUs and SMUs, 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.

### ***Discussion***

Our retrospective analysis of Chum Salmon CUs on the Inner South Coast and West Coast of Vancouver Island indicates that 25th and 75th percentile benchmarks are generally higher than Ricker-based benchmarks adopted under the Wild Salmon Policy, and are therefore a precautionary choice in data-limited situations. The few exceptions in our analyses were for Inner South Coast CUs, 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%  $S_{MSY}$ ). When using percentile benchmarks green status occurred in only 14% of years for CUs (9% for SMUs), but green status occurred in 79% of years for CUs (94% for SMUs) when Ricker benchmarks were used. 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 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 98% of CU-year combinations. However, benchmarks derived from the hierarchical Ricker model were more certain than those from the standard model in cases where productivity was similar across associated population units (CUs and SMUs). 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.

## Objective 2b. Evaluating benchmarks using simulation analyses

### Methods

We modified the simulation model developed in year 1 of the PSC funded project to evaluate performance of percentile-based benchmarks. That model was adapted from Holt and Folkes (2015) to evaluate data-limited benchmarks, described briefly here. Performance was measured as deviations between estimated lower benchmarks (25<sup>th</sup> percentile of observed spawner abundances and estimated  $S_{\text{gen}}$ ) and “true” lower benchmarks (“true”  $S_{\text{gen}}$ ). 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. 14). 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. A full description of model modifications from Holt and Folkes (2015) and parameterization is available in Appendix D. We made several revisions to the model from the first year of the project.

- A back-transformation bias correction for log-normally distributed variables was included with removed a positive bias in true and observed abundances.
- Sensitivity analyses on harvest rates were applied over the entire time-series of the simulation instead of only over the initialization period. This assumption more clearly reflects impacts of harvesting at those levels over the long-term, and reflects continued higher exploitation on some CUs (Georgia Strait) relative to others (Loughborough).
- Model results were evaluated in the context of West Coast Vancouver Island and Inner South Coast CUs. Specifically, performance of  $S_{25\text{th}}$  and  $S_{\text{gen}}$  benchmarks was evaluated for the CUs and stock management Units in those areas given estimates productivities and harvest rates. Due to the first two revisions, a re-examination of Inner South Coast CUs relative to the simulation modelling results was warranted.
- For the CUs for which our results suggested that the  $S_{25\text{th}}$  percentile benchmark was not precautionary compared with “true” estimates of the lower benchmark (“true”  $S_{\text{gen}}$ ), we further evaluated performance of alternative percentiles for lower benchmark, ranging from 30<sup>th</sup>-50<sup>th</sup> percentile in increments of 5%.

### *Sensitivity analyses*

Similar to analyses performed in the first year of the analyses, to assess the strength and direction of effects of input parameters on benchmark performance, we varied each input parameters individually while all others were held constant in a sensitivity analysis (Appendix D, Table D1). These sensitivity analyses did not assess the sensitivity of performance to interactions among input variables. 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).

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. We further evaluated the impacts of variability in productivity and harvest rates on assessment errors when estimating Ricker stock-recruitment parameters (which are used to derive data-rich benchmarks) and on resulting contrast in observed spawner data.

Data contrast has been used by Alaska Department and Fish and Game to identify the level of precaution (i.e., percentile of observed abundances) that should be afforded when identifying sustainable escapement goals (Clark et al. 2014). In particular, when the contrast in observed spawner time-series is high (maximum escapement/minimum escapement >8), they recommend a range covering the 15<sup>th</sup> to 60<sup>th</sup> percentiles of observed escapement for the sustainable escapement goal. When contrast is low, they recommend a wider range with a lower bottom limit (5<sup>th</sup> -65<sup>th</sup> percentiles). However, contrast in time-series will depend in part on historical harvest rates and productivity. We evaluated the impact of variability in harvest rates and productivity on data contrast in our simulation model, to assess covariation in those variables.

Given large uncertainties in underlying data and concerns about potential impacts of observation errors and biases in spawner numbers (e.g., due to incomplete sampling), we also explored bivariate sensitivity analyses of magnitude and bias of observation errors on benchmark performance.

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 analyses on productivity and harvest rates.

### ***Time-varying productivity***

We also evaluated impacts of temporal variability in productivity on benchmark performance, in the form of a step-like regime shift from a predominantly high productivity to a predominantly low productivity regime. Similar to the year 1 of the project, we evaluated one method to account for that variability by truncating the data used to estimate benchmarks to either the most recent low-productivity period to capture current conditions, or the historic high-productivity period to avoid a shifting baseline. 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 yet been applied to other species or regions, where the data required for assessing changes in productivity are often lacking. In contrast to the analysis applied in year 1 of the project, the model for evaluating benchmarks under temporal variability in productivity included the revisions in the bullets above.

## Results

Simulation model outputs for an example CU are presented in Fig. 15. Harvest rates during the initialization period were drawn at random from the historical time-series of exploitation rates for that CU, and the productivity parameter was estimated from the historical data using a hierarchical Ricker model (from Objective 1). Mean percent error between estimated and “true” benchmarks was generally greater than zero, especially for percentile benchmarks (Fig 15, right panel). For this CU, percentile benchmarks tended to be precautionary (i.e., deviations between 25<sup>th</sup> percentile benchmark and true  $S_{\text{gen}}$  were positive), whereas the stock-recruitment benchmarks tended to be negatively biased (i.e., deviations between estimate of  $S_{\text{gen}}$  and true value were negative). Estimates of stock-recruitment benchmarks differed from the “true” values because estimates were based on observed data (black line in Fig 15, left panel, and solid dots in Fig. 15, middle panel) rather than “true” data (grey line in Fig 15, left panel, and hollow dots in Fig. 15, middle panel). The assessed stock-recruitment model (black curve, Fig. 15, middle panel) differed from the “true” underlying model (grey curve, Fig. 15, middle panel) due to those errors in spawner abundance and time-series biases (Walters and Martell 2004).

Similar to results from Year 1 of the project, we found that 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. 16a 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). 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,  $S_{25\text{th}}$ , harvest rates had a strong 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 (not shown).

The global sensitivity analyses showed similar patterns as the univariate and bivariate sensitivity analyses, and similar to results from Year 1 of the project despite model changes. The mean elemental effects (magnitude of sensitivity, x-axis of Fig. 17) were greatest for productivity for both  $S_{25\text{th}}$  and  $S_{\text{gen}}$  benchmarks. Harvest rates were secondarily important for the  $S_{25\text{th}}$  benchmark (Fig. 17b). Parameters that ranked high on the standard deviation in elemental effect (y-axis of Fig. 17, e.g., observation errors in spawner) were influential for benchmark performance only in combination with other input parameters.

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 are precautionary (i.e., estimated benchmarks are equal to or higher than “true”  $S_{\text{gen}}$  lower benchmark) (Fig. 18, top left portion of panels; productivities are depicted as  $\log_e(\alpha)$ , Eqn. 2). At low productivity and high harvest rates, neither benchmark is precautionary (Fig. 18 bottom right portion of panels), and this true for  $S_{\text{gen}}$  even at low harvest rates (Fig. 18b, bottom left corner). These results differ from results from the 1<sup>st</sup> year of the project due to model changes described above.

When we superimposed CU-specific productivities and harvest rates for Inner South Coast Chum Salmon, the  $S_{25th}$  lower benchmarks were near the true  $S_{gen}$  benchmark for all CUs except Upper Knight, UK (Fig. 18a, symbols lie near the zero contour line, except UK). The  $S_{25th}$  benchmarks for 4 CUs was slightly (~25%) lower than true values (SCS= Southern Coast Streams, HSBI=Howe Sound and Burrard Inlet, NEVI= Northeast Vancouver Island, and GS=Georgia Strait). For Upper Knight, the  $S_{25th}$  benchmark was 100% greater than the true  $S_{gen}$ . We also evaluated CU-specific performance of alternative percentile benchmarks ranging for 30<sup>th</sup> – 50<sup>th</sup> in increments of 5%. We found that 50th percentile benchmark tended to higher than true estimate  $S_{gen}$  for all ISC CUs (Fig. 19; performance of  $S_{gen}$  is same as in Fig 18b, but is shown here for comparison). Plots of performance of remaining percentiles benchmarks are provided in the Appendix E.

For the West Coast of Vancouver Island, the  $S_{25th}$  benchmarks were equal to or higher than the true  $S_{gen}$  benchmarks for both CUs (SWVI =Southwest Vancouver Island and NWVI=Northwest Vancouver Island), and all five stock management units within SWVI (Fig. 20a). However, uncertainty bounds in productivity crossed the zero contour line for SWVI and 4 of the 5 stock management units within SWVI. In contrast, estimates of the  $S_{gen}$  benchmark were below the true lower benchmark (i.e., below the zero contour line) for SWVI and 3 stock management units within SWVI (Fig. 20b). Uncertainties bounds crossed the zero contour line for all stock management units and CUs except NWVI, which had estimates of  $S_{gen}$  that were higher than the true value.

For the upper benchmarks, for the Inner South Coast CUs, the  $S_{75th}$  benchmarks were higher than the true 80%  $S_{MSY}$  upper benchmark, but the uncertainty bounds for two CUs crossed the zero contour line (Fig. 21a). The estimates of 80% of  $S_{MSY}$  were below the true benchmark values for all Inner South Coast CUs (Fig. 21b). Similar trends were observed for the West Coast of Vancouver Island CUs. The  $S_{75th}$  benchmark tended to be higher than the true upper benchmark, but the estimated 80%  $S_{MSY}$  value tend to be lower than true values (Fig 22).

We found similar patterns in deviations in estimated Ricker parameters from the true values, as for  $S_{gen}$  deviations, though the direction of effects varied (Fig. 23). The estimated Ricker  $\log_e(\alpha)$  parameter (productivity parameter) tended to be higher and estimated Ricker  $b$  parameter (carrying capacity) tended to be lower than the true values at low productivity and high harvest rates (Fig. 23, bottom right corner of both panels). The opposite occurred at high productivity and low harvest rates.

In addition, we found that contrast in observed time-series of spawner abundances (maximum escapement/minimum escapement) was minimized at low productivity and high harvest rates and maximized at high productivity and low harvest rates, ranging from 2-20) (Fig. 24).

The effects of the magnitude of observation errors in spawner abundances were small compared with the effects of biases in spawner abundances (Fig. 25), but both were smaller than the effects of productivity and harvest rates (Fig. 25 compared with Fig. 18).

Our model assumed spawner abundances at equilibrium,  $S_{eq}$ , remained constant as productivity varied in sensitivity analyses (as in Holt and Bradford 2011). When we considered an alternate assumption where  $S_{max}$  remained constant, but  $S_{eq}$  declined as productivity declined, we found

similar patterns in the results (within ~10% MPE). This alternate assumption represents a scenario of simultaneous declines in capacity and productivity.

### ***Time-varying productivity***

As expected, regime shifts from high to low productivity were associated with increases in “true” (i.e., deterministic)  $S_{\text{gen}}$ , 80% of  $S_{\text{MSY}}$ , and  $S_{\text{max}}$  when  $S_{\text{eq}}$  was assumed constant (Fig. 26a-e). In contrast, when  $S_{\text{max}}$  was assumed constant, declines in productivity were associated with declines in “true” 80% of  $S_{\text{MSY}}$  and increases in “true”  $S_{\text{gen}}$  and  $S_{\text{eq}}$  (Fig. 26f-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. 27a) and higher estimates of  $S_{\text{gen}}$  (i.e., more precautionary) (Fig. 27c) under constant  $S_{\text{eq}}$ , compared to when the historical period was used. However, median estimates of  $S_{\text{gen}}$  were below the “true” value regardless of choice of data for inclusion (dashed line Fig. 27c), though the confidence intervals covered the “true” value in all three scenarios. The upper benchmark, 80% of  $S_{\text{MSY}}$ , did not change consistently with data truncation (Fig. 27e). Both  $S_{25\text{th}}$  and  $S_{75\text{th}}$  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”  $S_{\text{gen}}$  and 80% of  $S_{\text{MSY}}$  benchmarks, respectively. Although  $S_{\text{gen}}$  became more precautionary as data were truncated to the recent period and percentile benchmarks became less precautionary, percentile benchmarks were still consistently greater than true values (mean percent errors for  $S_{25\text{th}}$  and  $S_{75\text{th}}$  were  $\gg$  zero) (Fig. 28a). These results are similar to those presented in Year 1 of the project.

Similar patterns were observed under the assumption of constant  $S_{\text{max}}$  with an abrupt decline in productivity and  $S_{\text{eq}}$  (Fig. 28b and 29), 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. 29a), though the estimate of  $S_{\text{eq}}$  did decline (Fig. 29b). Confounding between estimates of productivity and  $S_{\text{eq}}$  results in a relatively low (instead of high) value for  $S_{\text{gen}}$  when only recent data are used (Fig. 29c).

## **Discussion**

We found that performance of percentile-based benchmarks was more sensitive to uncertainties in productivity and variability in harvest rates than to other model parameters, including observation errors in spawner abundances, catch, and age-at-maturity. Both  $S_{25\text{th}}$  and estimates of  $S_{\text{gen}}$  tend to be below the “true” lower benchmarks (“true”  $S_{\text{gen}}$ ) when harvest on unproductive CUs is high. As CUs are depleted, the time-series of observed abundances are dominated by low values, ratcheting the  $S_{25\text{th}}$  benchmark downward over time. The opposite occurs for highly productive CUs with low harvest where time-series are dominated by high abundances, pushing  $S_{25\text{th}}$  benchmark upwards.

Estimates of  $S_{\text{gen}}$  tend to be negatively biased when productivity is low due to time-series biases on stock-recruitment parameters. Time-series biases occur when the independent variable in stock-recruitment relationship (spawner abundances) depends on the recruitment (dependent

variable) in the previous generation, and are well documented for salmon populations (Walters and Martell 2004). The lack of independence between spawners and recruitment results in over estimates of productivity and under estimates in carrying capacity (e.g., as documented for Skeena River salmon in BC, Korman and English 2012), effects which are accentuated at low productivity (Korman et al. 1995) (as shown on the contours in Fig. 18). These parameter biases results in underestimates of  $S_{\text{gen}}$  (Fig. 30). Although state-space versions of stock-recruitment model that account for uncertainty in spawner abundances have been proposed as a way to address time-series biases, the performance of these methods against standard stock-recruitment models has been equivocal (Su and Peterman 2012). Alternatively, hierarchical models such as the models presented in Objective 1 have been suggested as a method to reduce these biases (Korman and English 2012), but a thorough evaluation of those methods under different scenarios of productivity and data quality is lacking. Further work evaluating impacts of hierarchical formulations of stock-recruitment models (and other Bayesian models with informative priors) on time-series biases and benchmark performance is warranted.

For the CUs we evaluated on the West coast of Vancouver Island, the percentile benchmarks tended to be more precautionary than the true  $S_{\text{gen}}$  benchmarks, but this was not the case for Inner South Coast because of low productivity and high harvest rates of several CUs in that region. We found that benchmarks based on the 50<sup>th</sup> percentile of observed spawner abundances,  $S_{50\text{th}}$  tended to be more precautionary than true  $S_{\text{gen}}$  benchmark (including all CUs evaluated here), except when productivities were between 1-1.2 and harvest rates were  $\geq 0.5$ , productivities were between 0.8-1 and harvest rates were  $\geq 0.4$ , or productivity was  $< 0.8$  and harvest rates were  $\geq 0.2$ . Under those scenarios, higher percentiles would be warranted to match the level of precaution provided by “true”  $S_{\text{gen}}$  benchmarks.

Under scenarios of time-varying productivity, we found values of percentile benchmarks were highest (most precautionary) when only historical data on high-productivity regime was used to avoid shifting baseline during CU depletion. In contrast, the performance of  $S_{\text{gen}}$  benchmarks depended on whether  $S_{\text{eq}}$  or  $S_{\text{max}}$  remained constant as productivity changed. Overall, under a scenario of step-like declines in productivity, our results suggest that estimates of  $S_{\text{gen}}$  are least biased when the entire times is used to estimate benchmarks instead of either historical or recent years due to strong time-series biases and confounding between stock-recruitment parameters that occur when time-series are short.

### ***Comparison to Sustainable Escapement Goals developed by ADFG (Clark et al. 2014)***

Clark et al. (2014) outlined sustainable escapement goals based on percentiles of observed spawner abundances using 4 tiers:

- Tier 1 – for high escapement contrast (greater than 8) and at least moderate harvest rate, the central 50-percentile range (25<sup>th</sup> to 75<sup>th</sup> percentiles)
- Tier 2 – for medium escapement contrast (4 to 8) and at most low harvest rate, the 15<sup>th</sup> percentile to the 75<sup>th</sup> percentile
- Tier 3 – for medium escapement contrast (4 to 8), the central 70-percentile range (15<sup>th</sup> to 85<sup>th</sup> percentiles)
- Tier 4 – for low escapement contrast (less than 4), the 15<sup>th</sup> percentile to maximum observed escapement (100<sup>th</sup> percentile).

Our evaluation differed from that of Clark et al. (2014) because we were interested in evaluating percentile benchmarks in context of Wild Salmon Policy benchmarks, and not escapement goals based on MSY. The lower bounds of the sustainable escapement goals represent a lower bound of the escapement goals, which we focused our comparison against.

Similar to the results of Clark et al. (2014), we recommend higher percentiles be considered when harvest rates are high. In contrast to Clark et al. (2014), we found that contrast in escapement data was correlated with stock productivity, and we recommend higher percentiles under low productivity and low contrast, instead of low percentiles (as in Clark et al. (2014)). These different results may be explained in part by the smaller range in productivities considered by Clark et al. (2014) in their simulation evaluation of percentile-based escapement goals ( $\log_e(\alpha)=1-2$ ), compared with the range considered here ( $\log_e(\alpha)=0.5-2$ ).

### Objective 3: Recommendations on application of benchmarks

Our retrospective analysis shows that for Chum Salmon CUs examined here status determined using percentile-based benchmarks were consistently more precautionary than Ricker-based benchmarks, and are therefore may be a viable choice for use in sites where Ricker-based benchmarks cannot be calculated or are highly uncertain. However, we recommend caution when using percentile benchmarks when productivity is moderate-low ( $\log_e(\alpha) < 1.2$ ) and and/or harvest rates are moderate-high ( $>0.2$ ) based on results from our simulation modelling. In these cases, the performance of benchmarks depends on the combination of productivity and harvest rates. Even if  $S_{25th}$  benchmarks are higher than then estimates of the data-rich,  $S_{gen}$  benchmark,  $S_{25th}$  may be lower than the “true”  $S_{gen}$  benchmark resulting in possible overestimates of status. To avoid overestimates of status in these cases, benchmarks based higher percentiles of spawner time-series (e.g.  $S_{50th}$  instead of  $S_{25th}$  for the lower benchmark) may be warranted. Therefore, the choice of percentile for the lower benchmark should vary among CUs depending on productivity and harvest rates. When productivity and/or harvest rates are highly uncertain, a precautionary approach would be to adopt a relatively high percentile ( $S_{50th}$  = median of spawner time-series) as a lower benchmark. We note, however, that there is a level of productivity below which, and harvest rates above which,  $S_{50th}$  benchmark will no longer be precautionary (below zero contour line in Fig. 19a), and percentile-based benchmarks are not recommended. One caveat on the application of percentile-based benchmarks is that uncertainties in benchmarks are not provided.

Our simulation results found that for 4 CUs (Southern Coast Streams, Georgia Strait, Howe Sound/Burrard Inlet, and Northeast Vancouver Island),  $S_{25th}$  benchmarks were lower than the true underlying benchmarks and therefore may fail to detect conservation concerns. For these CUs,  $S_{50th}$  benchmarks were larger than true benchmarks, and therefore may better detect conservation concerns when they exist. These 4 CUs were characterized by recent declines in productivity (Southern Coastal Streams and Northeast Vancouver Island) or high harvest rates (Georgia Strait and Howe Sound/Burrard Inlet). For the remaining CUs on Inner South Coast and West Coast of Vancouver Island (including component stock management units), values of  $S_{25th}$  were equal to or greater than true underlying lower benchmark. However, uncertainties in productivity and variability in harvest rates for these CUs are high, and so status above  $S_{25th}$  may not necessarily protect CUs against conservation risks given that a large portion of the distributions are associated with enhanced risks.

These results can be applied to other CUs of Chum Salmon in BC, including Fraser River. Where productivity and harvest rates are known to be low,  $S_{25th}$  percentile benchmarks are recommended for lower benchmarks of biological status. Where productivity estimates and/or harvest rates are low or highly uncertain, a higher benchmark, such as  $S_{50th}$  may be warranted.

## General Discussion: Time-varying Productivity

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 Mallick 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 when regime-like shifts in productivity occur and percentile-based benchmarks are used, truncating data to a historical high-productivity period avoids the shifting baseline phenomenon that occurs when the entire time-series or only the most recent data are used. However, for stock-recruitment based benchmark,  $S_{\text{gen}}$ , performance is improved by using the entire time-series, as that benchmark is influenced more strongly by time-series biases and confounding between Ricker parameter which occur especially in short time-series.

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 time-varying productivity parameter analyses from section 1 of this report 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.
- Davis B. 2017. ChumGEM Model Testing and Sensitivity Analysis. Report submitted to the Pacific Salmon Commission Chum Technical Committee by DFO. 6 January, 2017.
- 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.
- 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.

Ryall P, Murray C, Palmero V, Bailey D, Chen D. 1999. Status of Clockwork Chum Salmon Stock and Review of the Clockwork Management Strategy. Canadian Stock Assessment Secretariat Res. Doc. 99/169.

Su Z, Peterman RM. 2012. Performance of a Bayesian state-space model of semelparous species for stock-recruitment data subject to measurement error. Ecological Modelling 224:76-89.

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.

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

	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	1.60	2.43	1.53	2.23	1.70	2.43	2.22	4.04	2.18	3.56
		0.92		1.02		1.05		1.14		1.19		1.34
$S_{\max}$	80,275	218,015	67,219	183,913	115,696	313,299	101,040	284,140	16,523	62,829	16,756	57,175
		43,303		41,050		68,614		63,165		9,410		9,804
$S_{\text{gen}}$	9,636	19,263	9,994	18,702	16,506	31,386	16,292	31,635	2,944	9,167	3,089	9,365
		2,131		4,292		7,927		10,155		1,485		1,736
80% $S_{\text{MSY}}$	10,372	20,370	11,711	20,687	18,503	33,262	19,494	34,620	4,581	11,619	4,578	12,270
		1,734		3,647		6,688		9,098		1,980		2,429
	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 $\alpha$	2.30	3.23	2.24	3.06	2.46	3.77	2.32	3.44	3.08	4.77	2.67	4.19
		1.59		1.66		1.62		1.64		2.05		1.97
$S_{\max}$	62,730	123,151	64,033	116,832	106,264	246,135	111,430	278,742	493,198	1,083,934	608,911	1,143,835
		43,696		44,811		69,179		73,740		301,072		336,236
$S_{\text{gen}}$	11,992	21,103	12,227	20,857	20,222	44,095	21,257	47,248	90,983	206,206	116,883	216,737
		8,095		8,440		12,203		13,489		44,872		53,905
80% $S_{\text{MSY}}$	18,401	27,404	18,194	27,109	33,348	60,030	33,484	62,854	186,802	303,619	203,327	313,297
		13,671		13,628		23,044		22,782		141,661		146,895
	Howe Sound to Burrard Inlet				North West Coast Vancouver Island				South West Coast Vancouver Island			
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 $\alpha$	2.63	3.75	2.47	3.49	2.50	3.75	2.51	3.76	2.81	4.55	2.78	4.56
		1.81		1.79		1.66		1.65		1.69		1.64
$S_{\max}$	511,173	1,657,672	559,155	1,837,905	62,398	108,771	62,597	110,640	344,426	731,855	348,571	746,969
		308,310		333,798		44,590		44,742		232,644		235,912
$S_{\text{gen}}$	97,554	310,845	107,571	344,097	11,995	18,845	11,997	18,841	65,109	124,899	66,202	125,382
		54,892		60,229		7,754		7,724		35,981		36,558
80% $S_{\text{MSY}}$	171,126	410,187	177,421	453,338	19,871	25,247	19,921	25,492	120,726	161,752	121,273	160,359
		119,131		120,094		15,779		15,671		99,282		99,754

**Table 1b** - Parameter and benchmark estimates and upper/lower credible interval bounds delineated as 2.5th and 97.5th posterior densities for most recent year

	Esperanza				Barkley				Clayoquot			
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 $\alpha$	3.53	5.55	3.47	5.48	1.95	3.02	2.09	3.24	2.28	3.59	2.43	3.72
		2.15		2.19		1.24		1.32		1.44		1.53
$S_{\max}$	39,262	67,922	39,663	65,859	157,925	377,062	147,015	339,015	80,019	168,738	74,678	147,162
		28,526		28,732		102,715		97,110		52,782		50,454
$S_{\text{gen}}$	6,874	12,879	6,993	12,604	28,400	47,412	27,393	45,047	15,300	25,683	14,344	24,327
		3,947		3,973		18,834		17,564		9,328		8,750
80% $S_{\text{MSY}}$	16,227	21,087	16,224	20,746	38,056	53,280	38,650	52,026	23,037	31,325	23,072	30,609
		13,322		13,390		23,582		26,743		17,396		18,105
	Kyuquot				Nootka							
Model	Standard		Hierarchical		Standard		Hierarchical					
Statistic	Estimate	UCL/LCL	Estimate	UCL/LCL	Estimate	UCL/LCL	Estimate	UCL/LCL				
Ricker $\alpha$	4.65	7.05	4.38	6.82	3.71	6.09	3.59	5.89				
		2.89		2.85		2.22		2.23				
$S_{\max}$	47,709	67,775	49,044	68,769	41,043	68,610	41,959	68,734				
		38,042		38,454		29,942		30,531				
$S_{\text{gen}}$	7,185	12,632	7,648	12,744	7,035	13,068	7,307	13,096				
		4,325		4,569		3,849		4,035				
80% $S_{\text{MSY}}$	22,836	27,065	22,950	27,153	17,465	23,059	17,596	22,875				
		20,086		20,019		14,489		14,395				

Table 2a. 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	10,372	11,711	54,350
Lower Benchmark	9,636	9,994	5,425
	North East Vancouver Island		
Upper Benchmark	18,503	19,494	75,136
Lower Benchmark	16,506	16,292	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
	North West Coast Vancouver island		
Upper Benchmark	19,871	19,921	73,650
Lower Benchmark	11,995	11,997	24,811
	South West Coast Vancouver Island		
Upper Benchmark	120,726	121,273	433,640
Lower Benchmark	65,109	66,202	204,065

Table 2b. 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>).

Method	Esperanza		
	Standard Ricker	Hierarchical Ricker	Percentile
Upper Benchmark	16,227	16,224	54,242
Lower Benchmark	6,874	6,993	25,390
	Barkley		
	Standard Ricker	Hierarchical Ricker	Percentile
Upper Benchmark	38,056	38,650	145,222
Lower Benchmark	28,400	27,393	48,106
	Clayoquot		
	Standard Ricker	Hierarchical Ricker	Percentile
Upper Benchmark	23,037	23,072	76,239
Lower Benchmark	15,300	14,344	34,656
	Kyuquot		
	Standard Ricker	Hierarchical Ricker	Percentile
Upper Benchmark	22,836	22,950	84,739
Lower Benchmark	7,185	7,648	36,590
	Nootka		
	Standard Ricker	Hierarchical Ricker	Percentile
Upper Benchmark	17,465	17,596	50,502
Lower Benchmark	7,035	7,307	24,654

Table 3a. Conservation status for each CU for the most recent year of analysis, 2012 for ISC, 2015 for WCVI. Statuses are calculated using all data available for the Ricker-based benchmarks, and use all escapement data 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
North West Coast Vancouver island	Red	Green	Green
South West Coast Vancouver Island	Green	Green	Green

Table 3b. Conservation status for each WCVI SMU for the most recent year of analysis, 2015. Statuses are calculated using all data available for the Ricker-based benchmarks, and use all escapement data for the percentile-based benchmark.

Conservation Unit	Percentile Status	Standard Ricker Status	Hierarchical Ricker Status
Barkley	Red	Amber	Amber
Clayoquot	Red	Green	Green
Nootka	Amber	Green	Green
Esperanza	Amber	Green	Green
Kyuquot	Amber	Green	Green

Table 4a. 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	0.14	0.19	0.96	0.97
Upper Knight	0.51	0.49	0.79	0.77
Loughborough	0.21	0.23	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
North West Coast Vancouver island	0.08	0.08	1.00	1.00
South West Coast Vancouver Island	0	0	1.00	1.00

Table 4b. Proportion of years where Ricker-based status and percentile-based status match, by WCVI SMU 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
Barkley	0	0	1.00	1.00
Clayoquot	0.10	0.10	1.00	1.00
Nootka	0.08	0.08	1.00	1.00
Esperanza	0.03	0.03	1.00	1.00
Kyuquot	0.13	0.13	1.00	1.00

## Figures

Fig. 1. Schematic of ChumGEM model for reconstruction return abundances of Chum Salmon in southern BC and Washington State (Cox and Rossi, unpublished presentation).

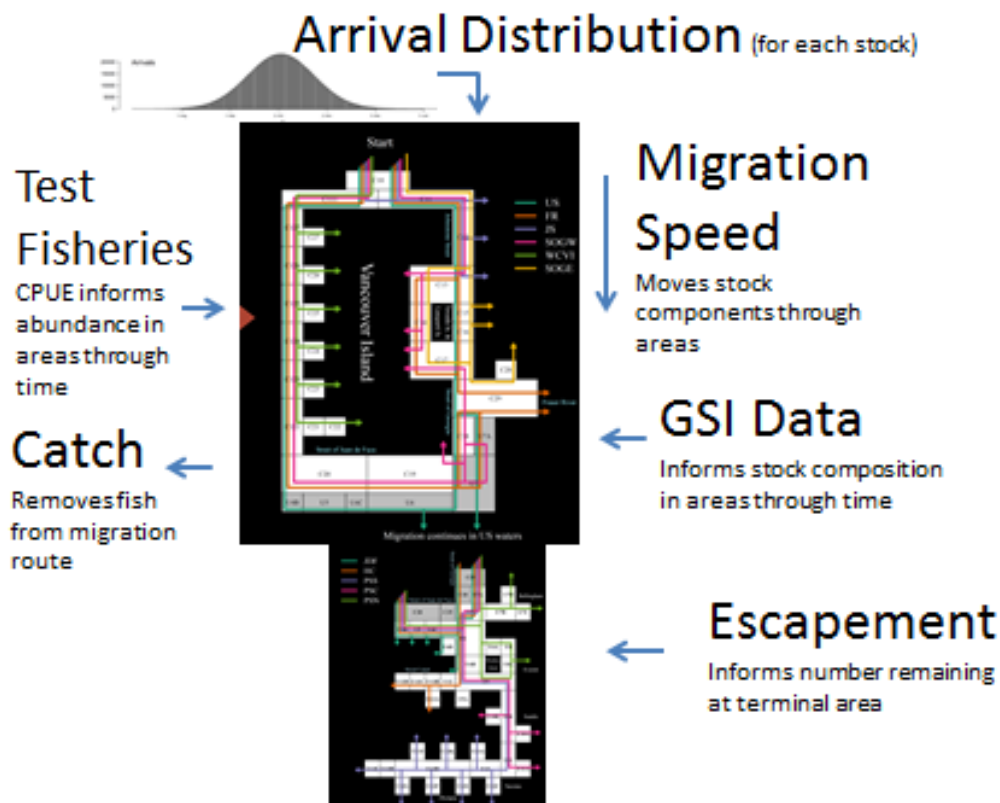


Fig. 2. Model estimates for Ricker  $\alpha$  (top row) and  $S_{max}$  (bottom row) across prior distributions (uniform and lognormal) for  $S_{max}$  and standard and hierarchical Ricker model structures (solid and dotted lines, respectively) for each CU in 2012. Circles indicate posterior medians,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

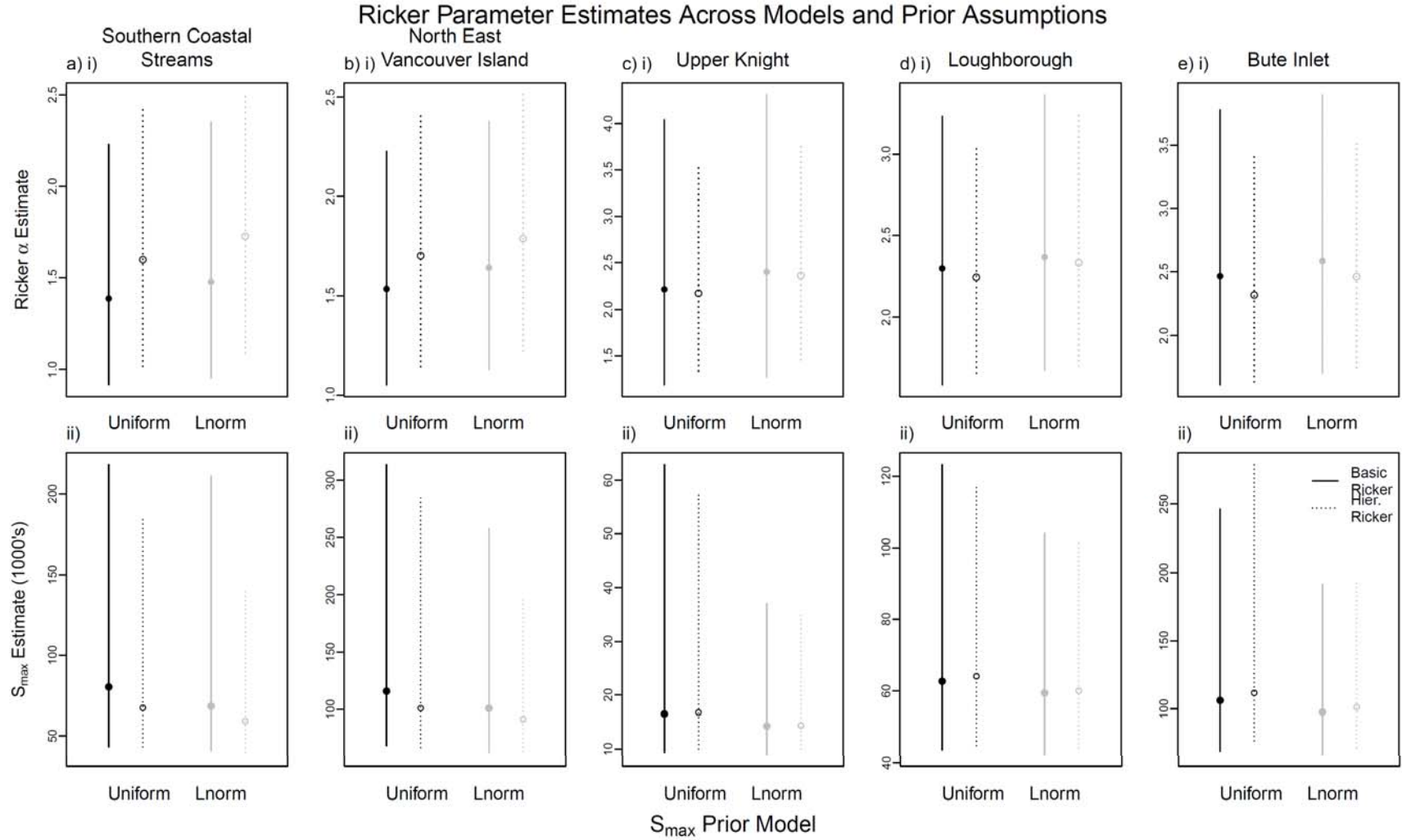


Fig. 2. cont.

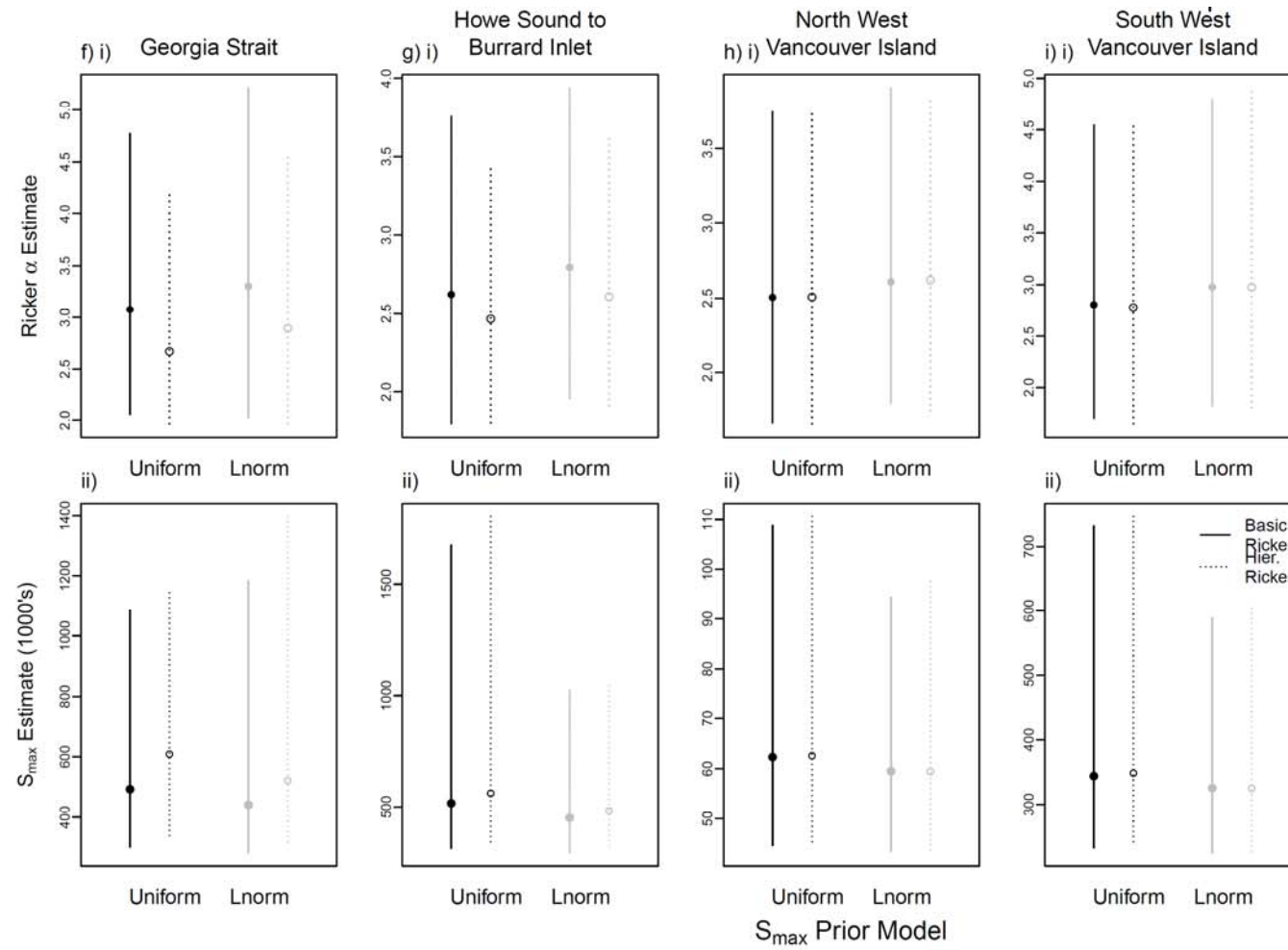


Fig. 3. Model estimates for Ricker  $\alpha$  (top row) and  $S_{max}$  (bottom row) across prior distributions (uniform and lognormal) for  $S_{max}$  and standard and hierarchical Ricker model structures (solid and dotted lines, respectively) for each SMU in 2012. Circles indicate posterior medians, and lines indicate 95% credible intervals of estimates.

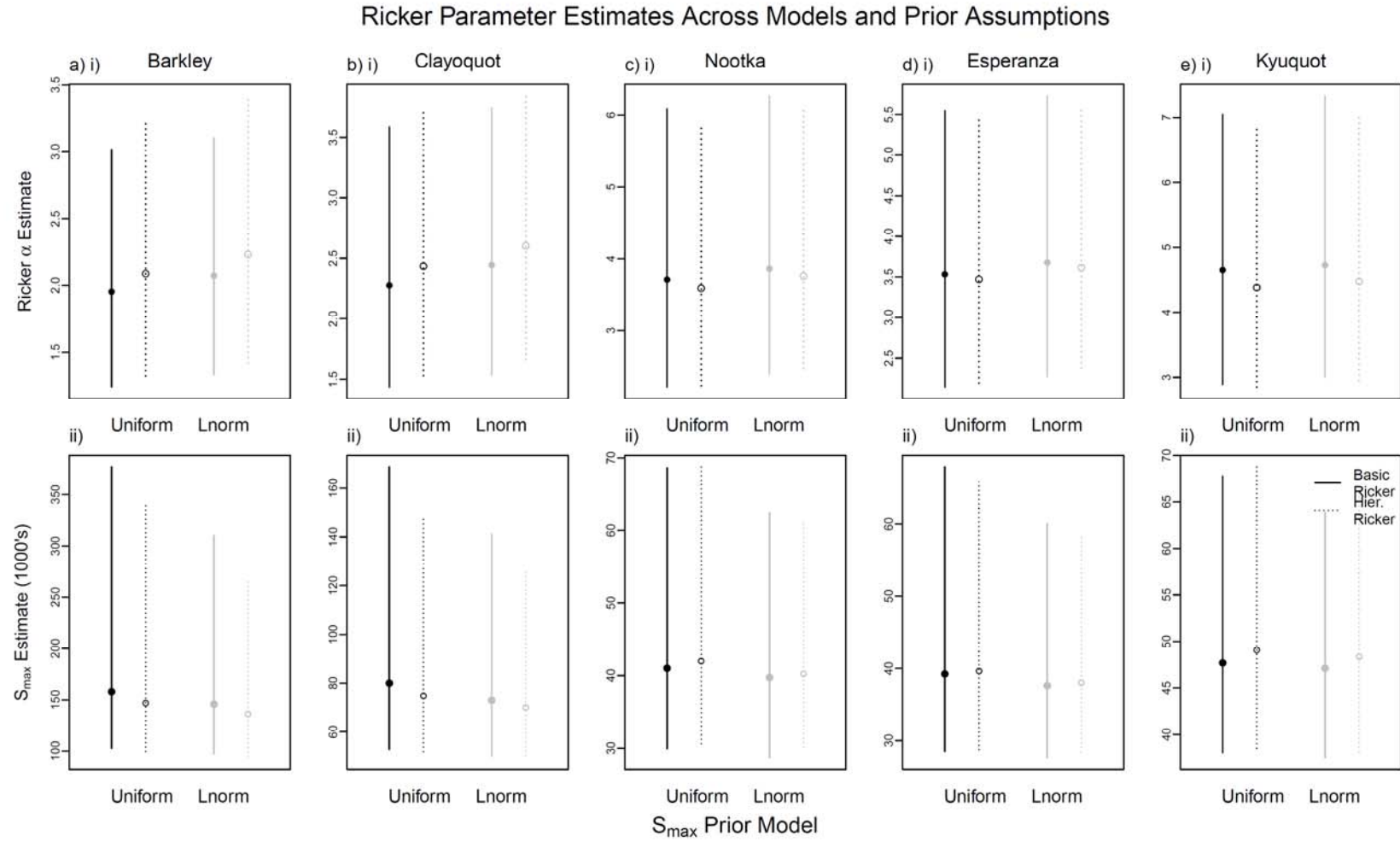


Fig. 4. Observed spawner-recruit data over time, with fitted Ricker curves and associated benchmarks for (i) the standard Bayesian Ricker model, and (ii) the Bayesian hierarchical Ricker model for CUs. Shaded regions indicate 95% credible intervals, delineated by 2.5th and 97.5th posterior densities. 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.

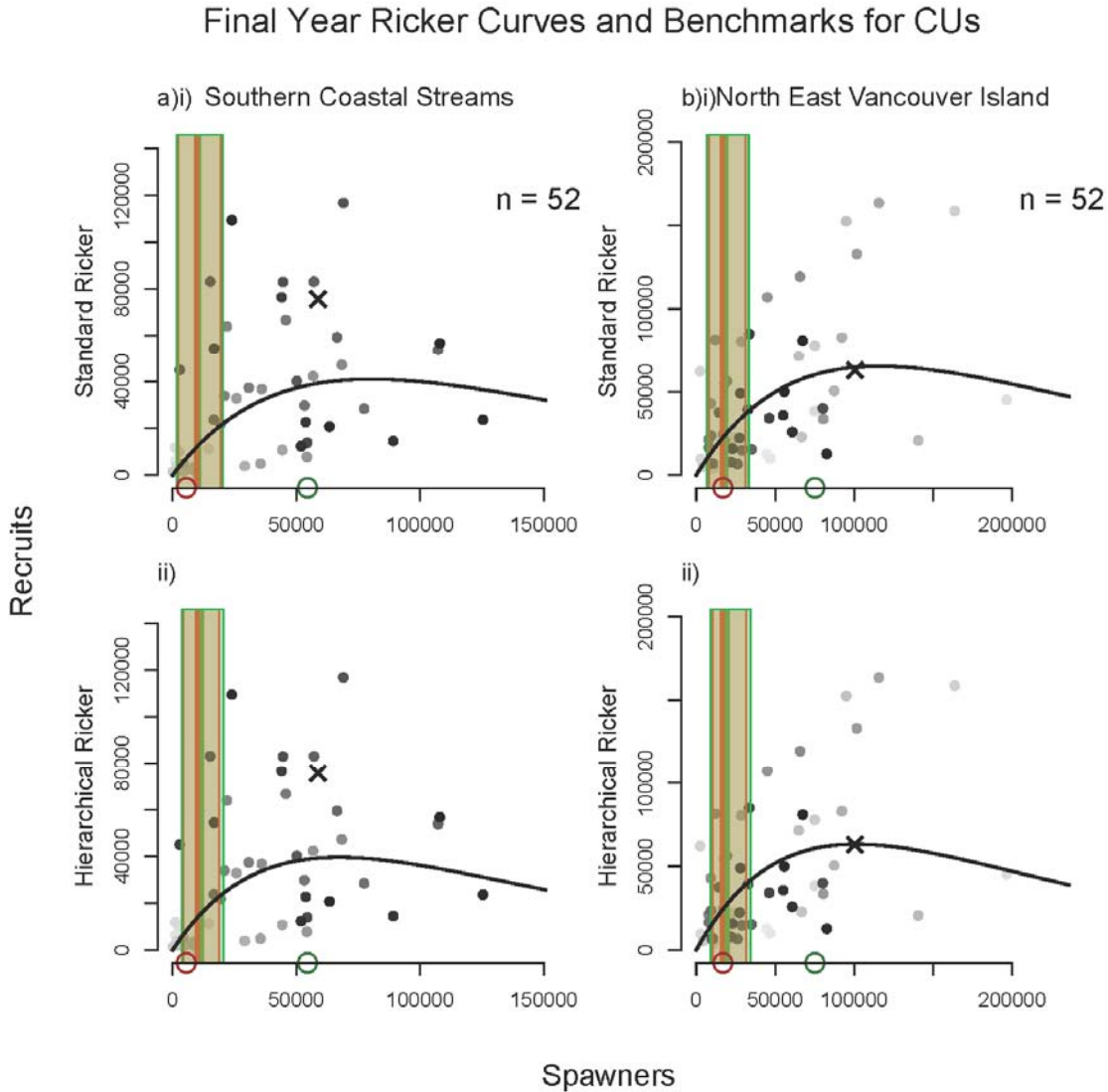


Fig. 4 cont.

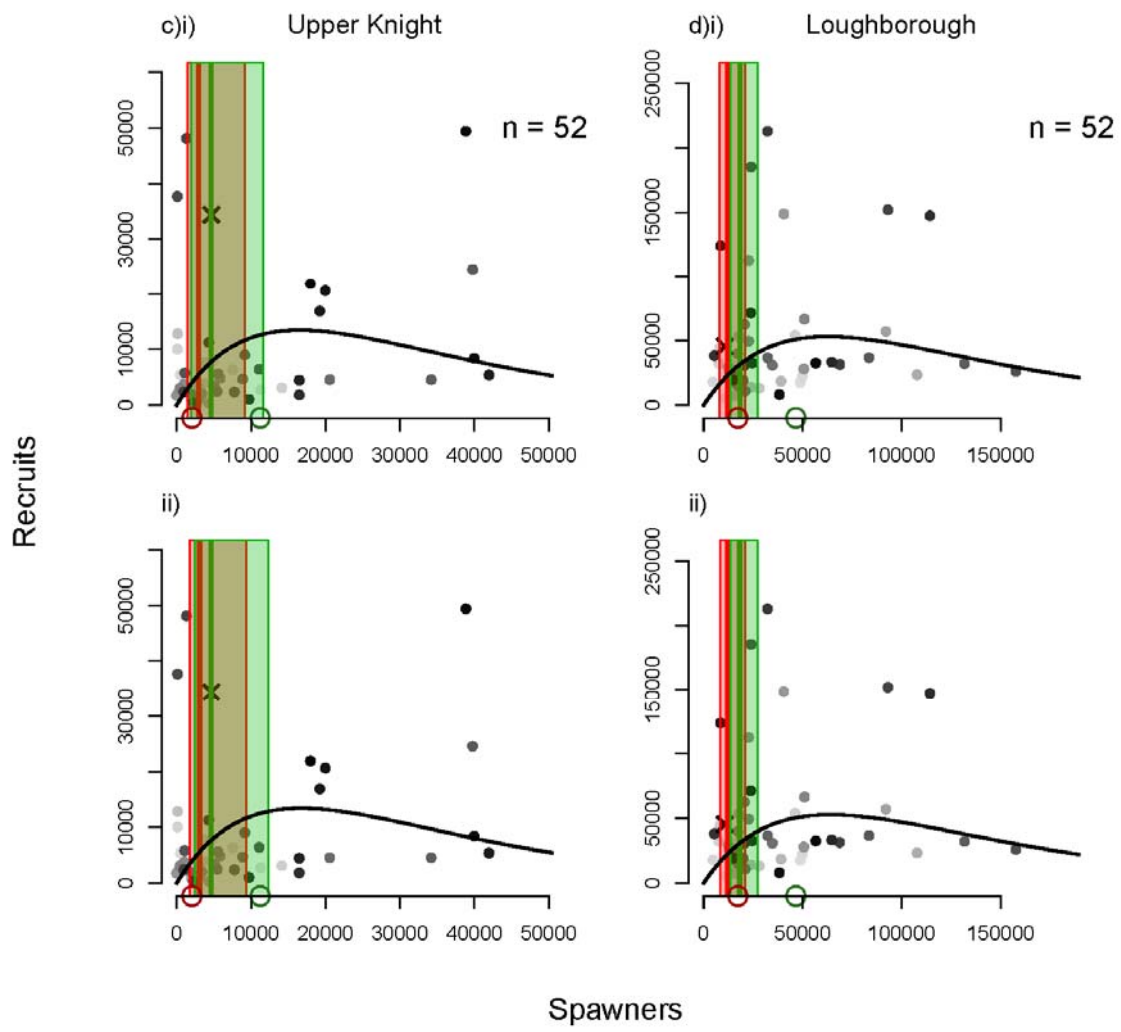


Fig. 4. cont.

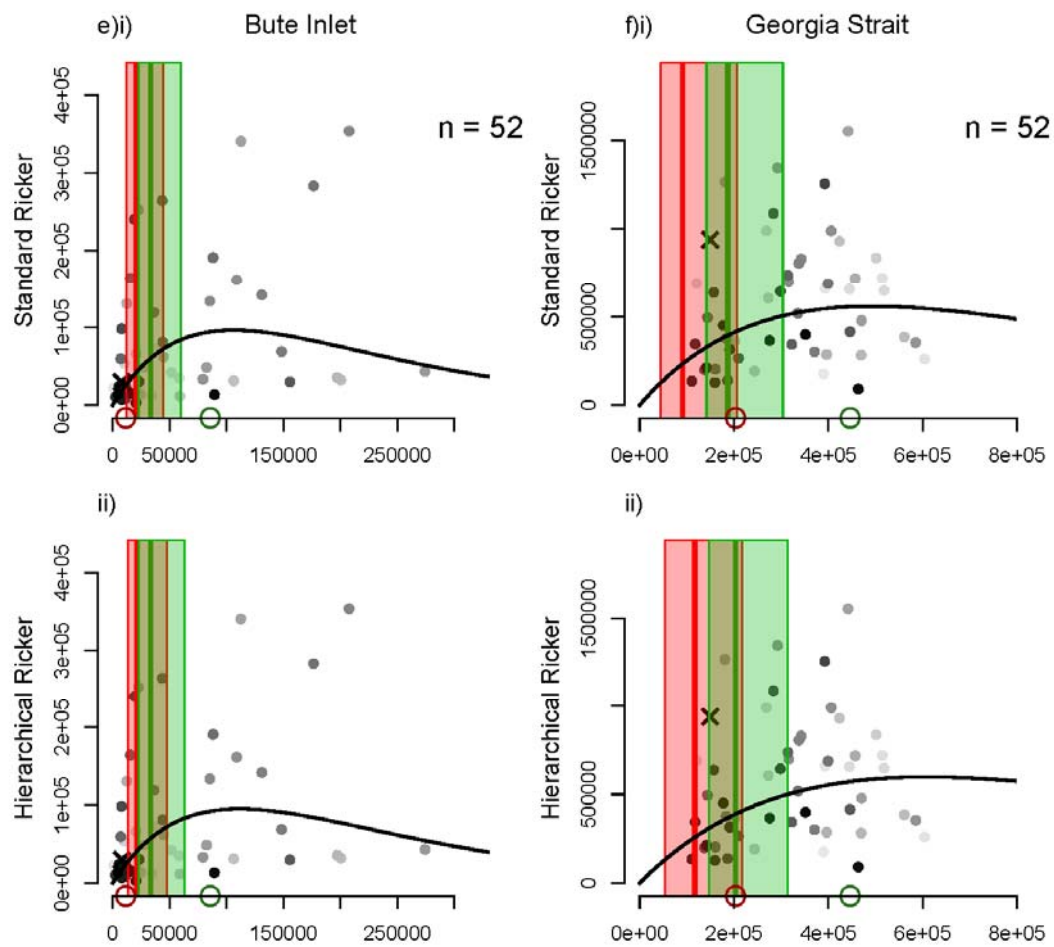


Fig. 4. cont.

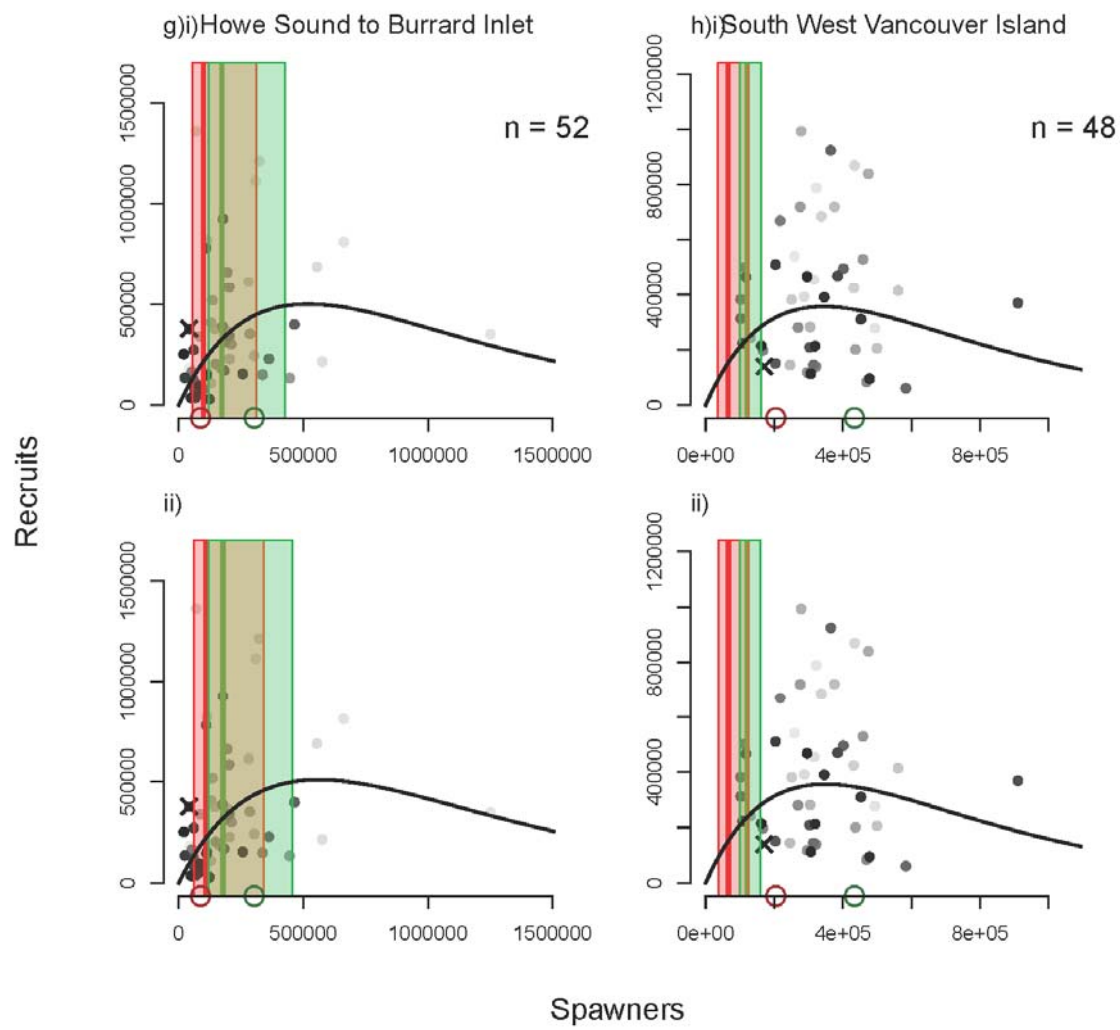


Fig. 4. cont.

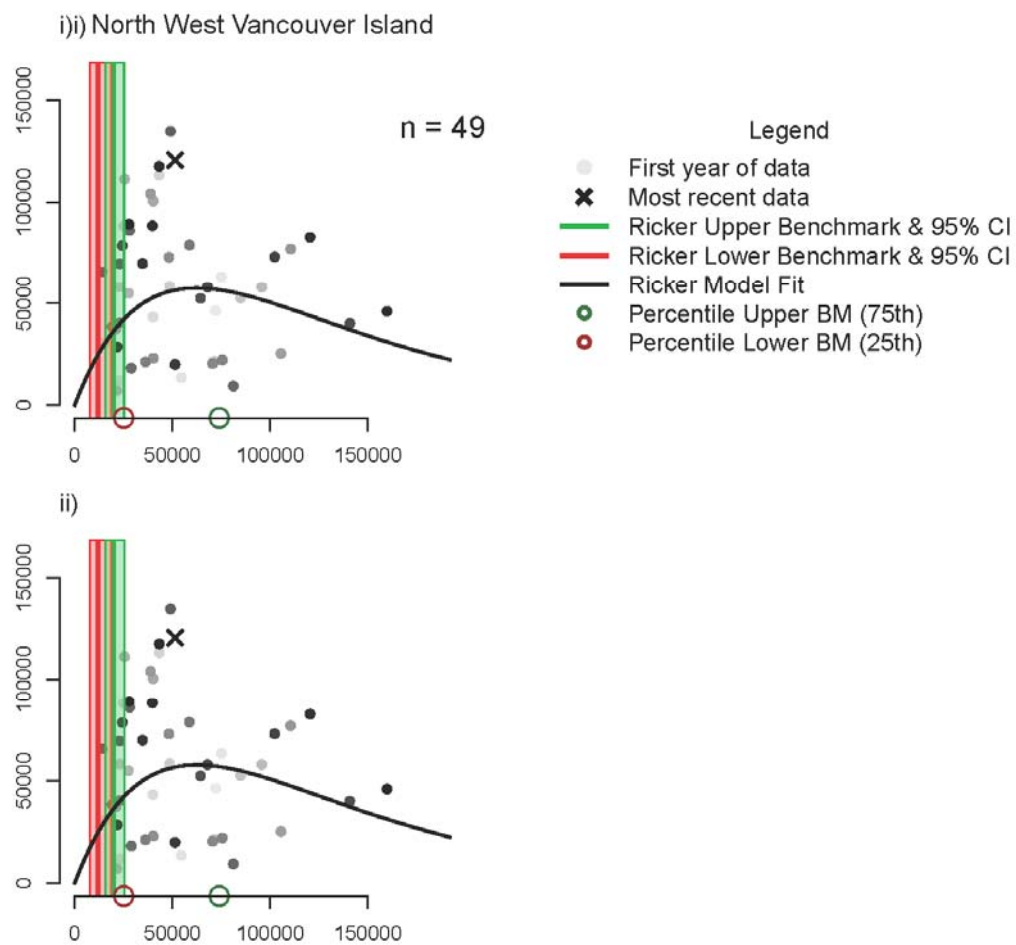


Fig. 5. Observed spawner-recruit data over time, with fitted Ricker curves and associated benchmarks for (i) the standard Bayesian Ricker model, and (ii) the Bayesian hierarchical Ricker model for SMUs. Shaded regions indicate 95% credible intervals, delineated by 2.5th and 97.5th posterior densities. 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.

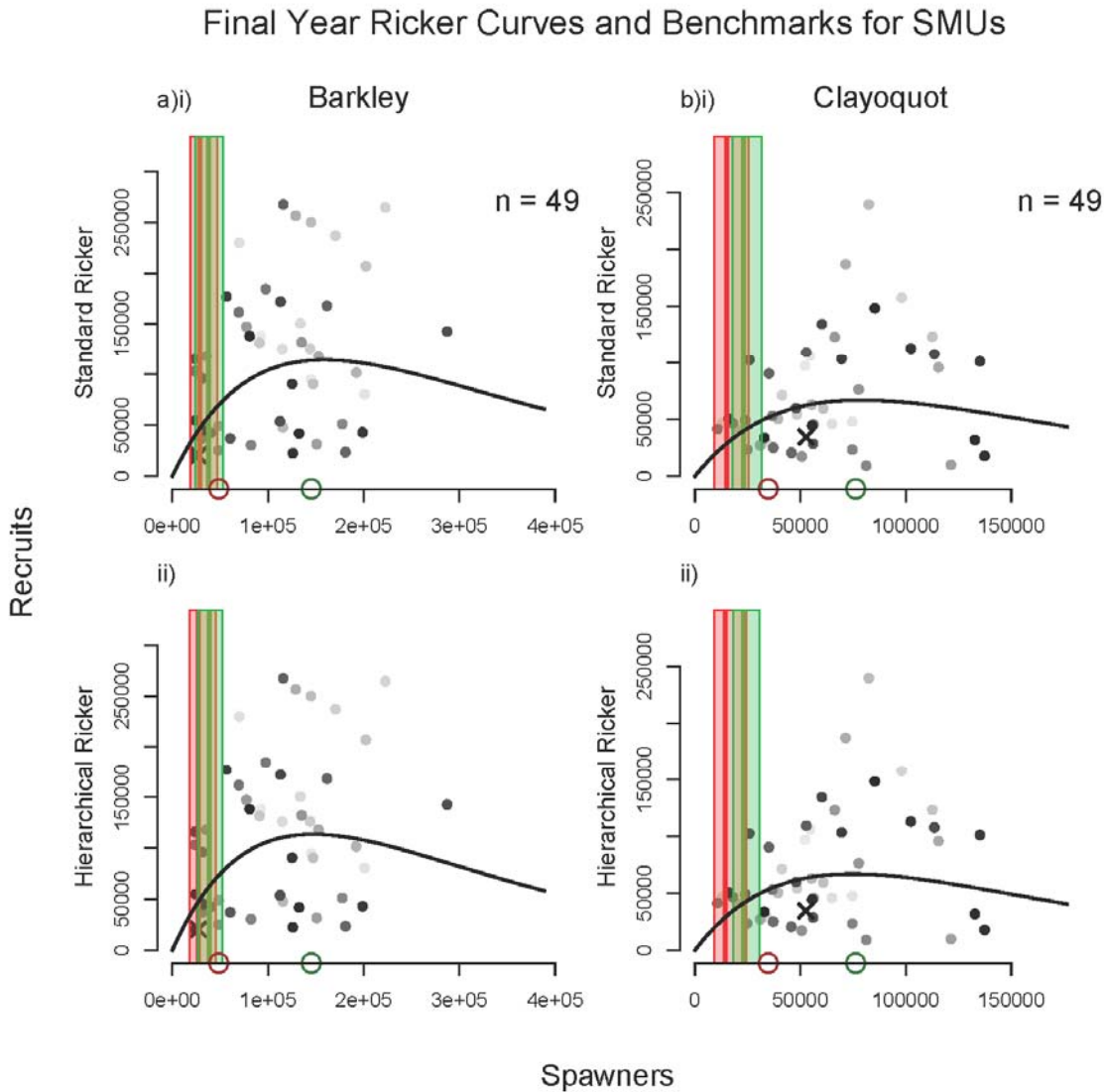


Fig. 5 cont.

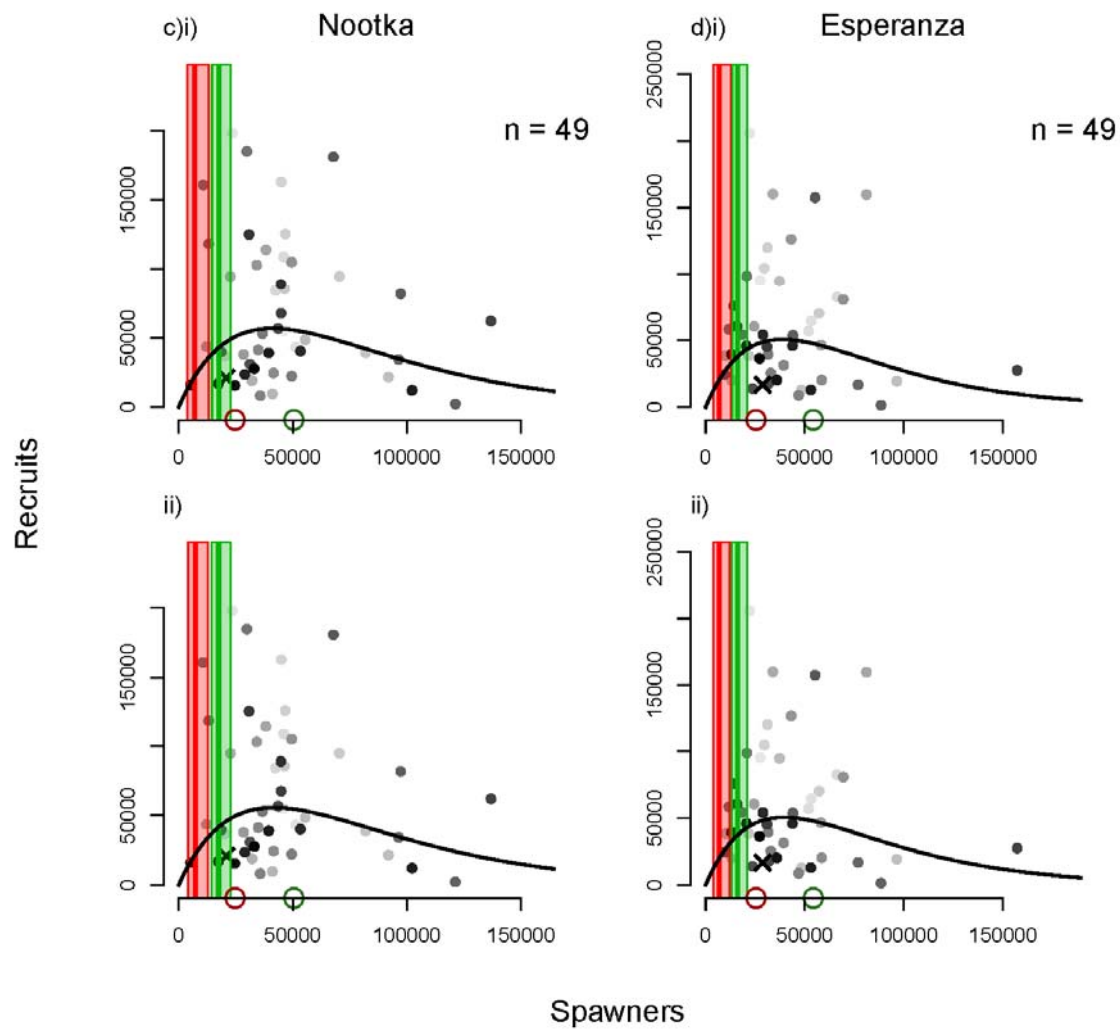


Fig. 5 cont.

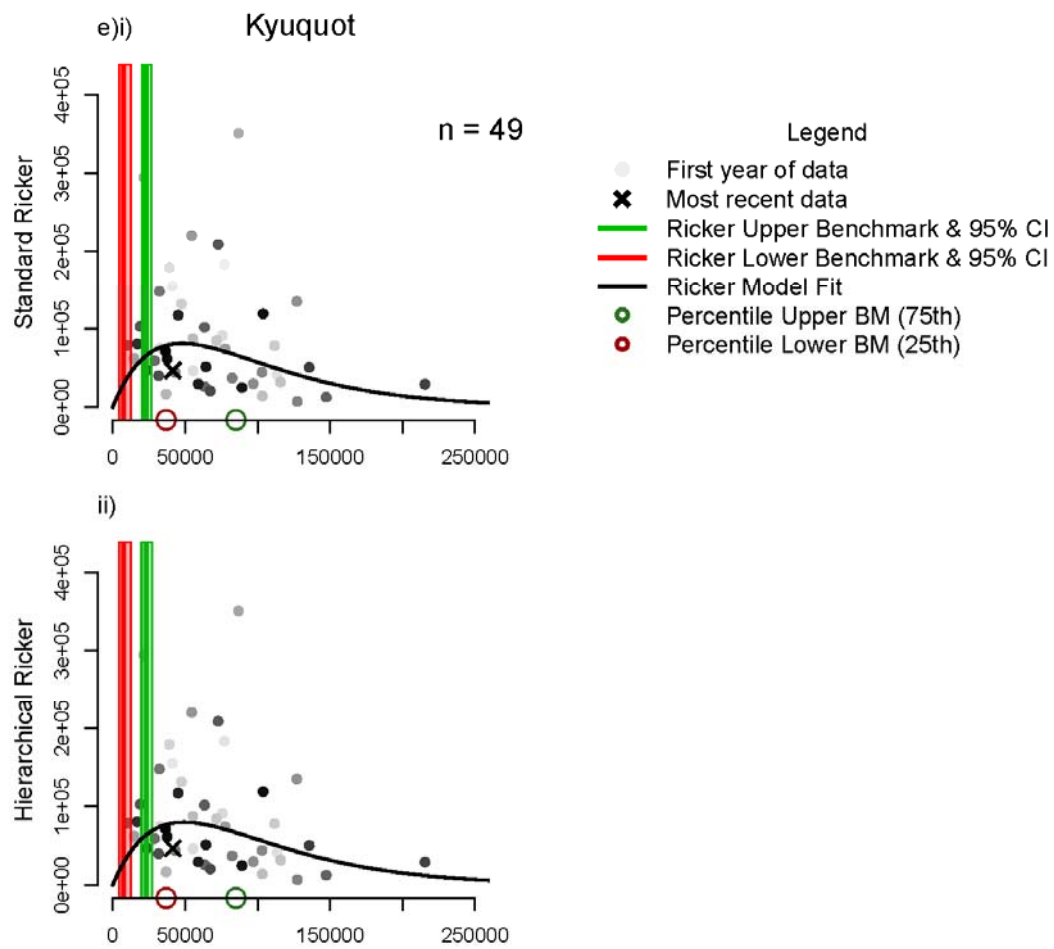


Fig. 6. Raw and generational average escapement over time for CUs, with retrospective conservation benchmarks overlaid for three benchmarks types: (i) standard Ricker model; (ii) hierarchical Ricker model; and (iii) percentile. Shaded regions indicate 95% credible intervals, delineated by 2.5<sup>th</sup> and 97.5<sup>th</sup> posterior densities. Retrospective benchmarks use all available data up to that year to estimate benchmark values.

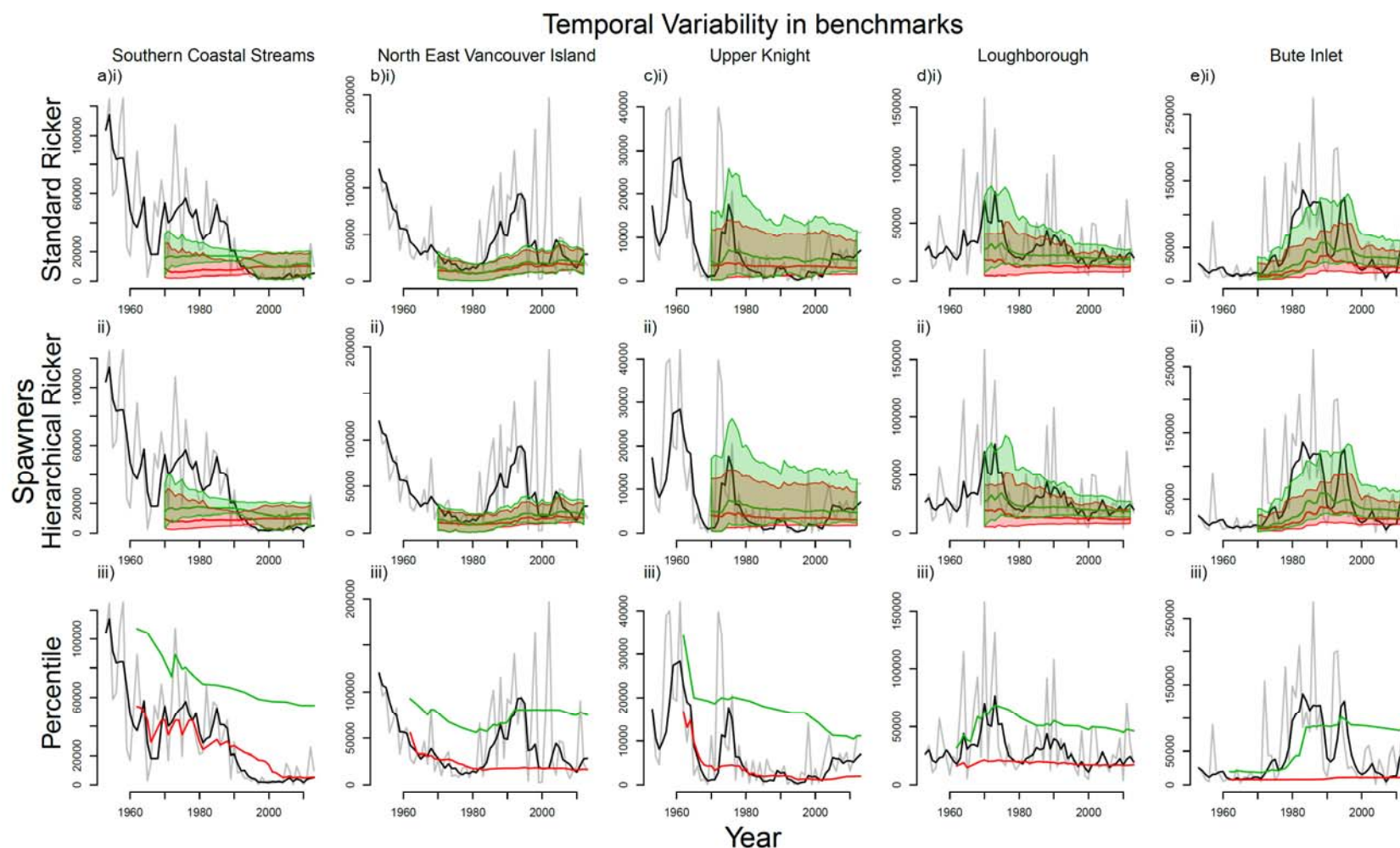


Fig. 6 cont.

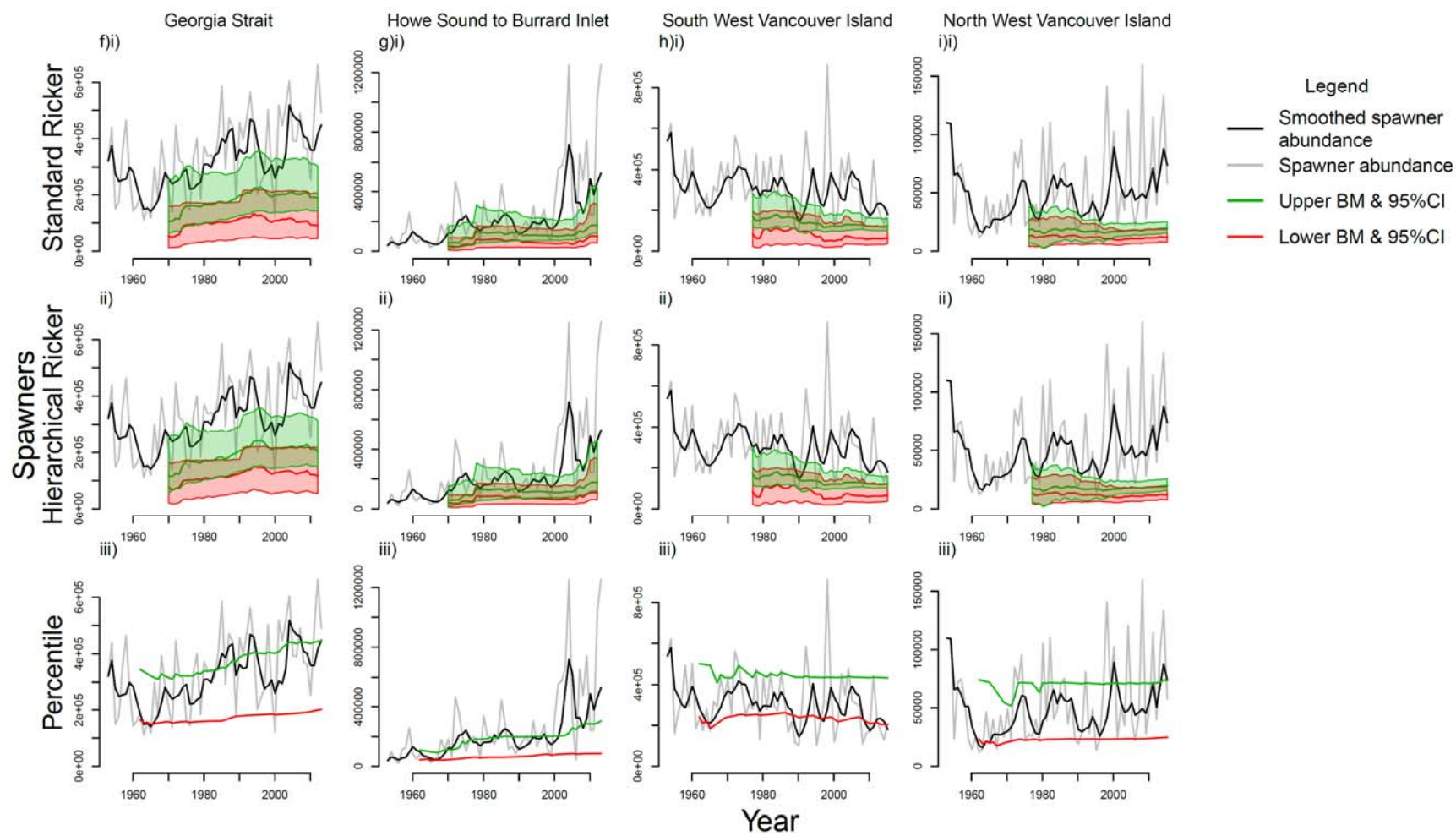


Fig. 7. Raw and generational average escapement over time for SMUs, with retrospective conservation benchmarks overlaid for three benchmarks types: (i) standard Ricker model; (ii) hierarchical Ricker model; and (iii) percentile. Shaded regions indicate 95% credible intervals, delineated by 2.5<sup>th</sup> and 97.5<sup>th</sup> posterior densities. Retrospective benchmarks use all available data up to that year to estimate benchmark values.

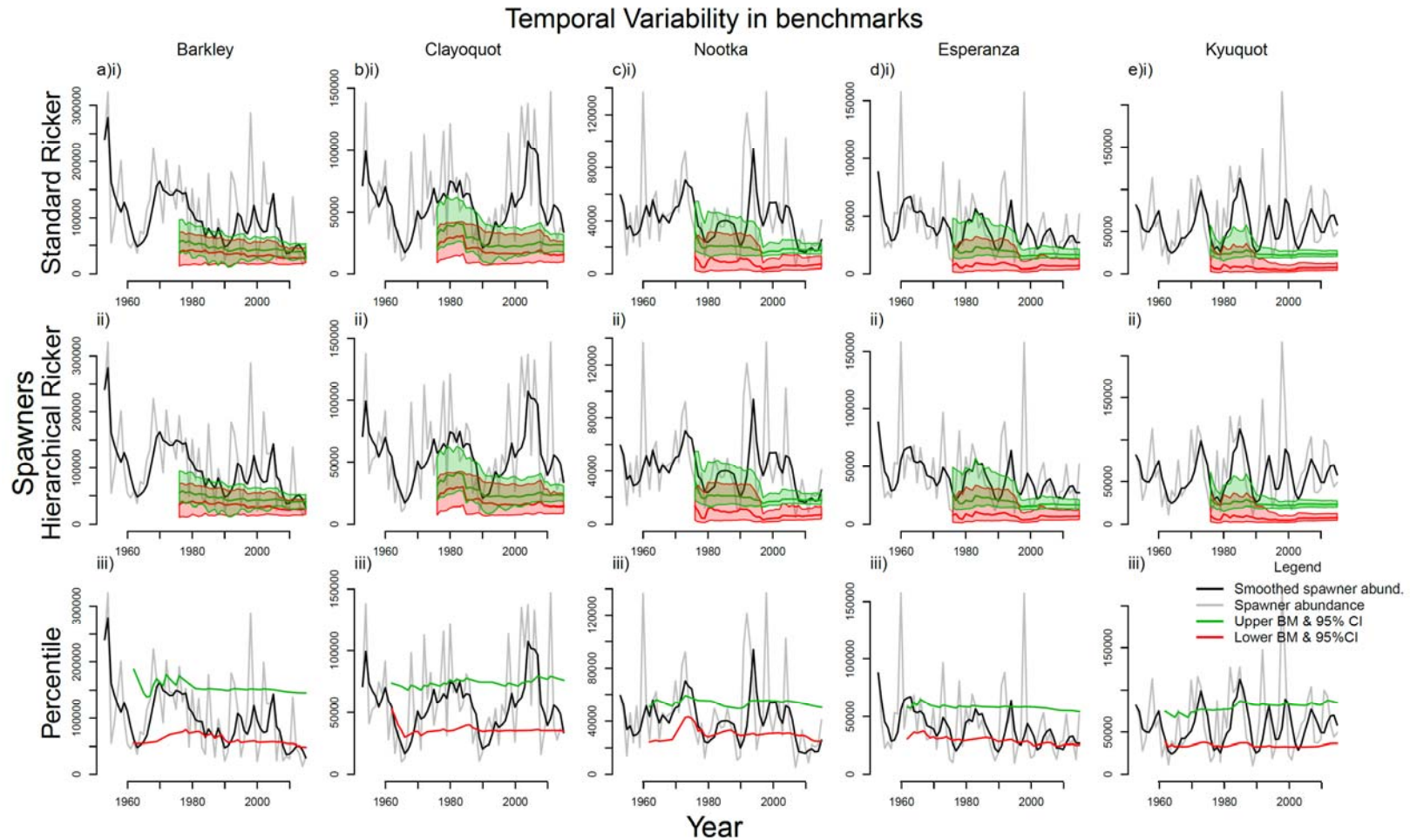


Fig. 8. Standardized raw and generational average escapements across CUs, with status indicated by coloured bars below. Transparent bars indicate upper and lower credible interval bounds, based on 2.5th and 97.5th percentiles of posterior distribution of estimated parameters. Gaps exist for Southern 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).

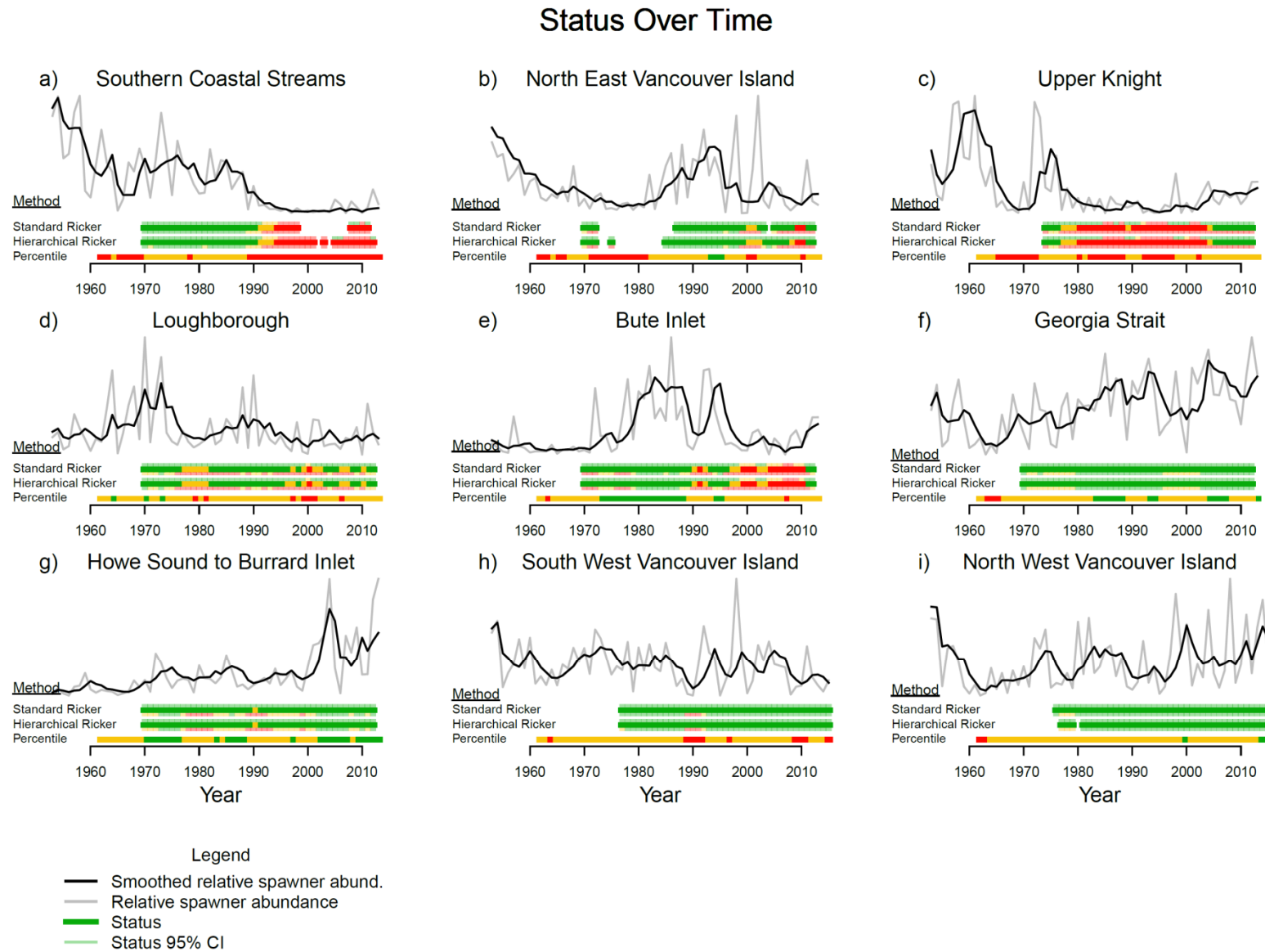


Fig. 9 Standardized raw and generational average escapements across SMUs, with status indicated by coloured bars below. Transparent bars indicate upper and lower credible interval bounds, based on 2.5th and 97.5th percentiles of posterior distribution of estimated parameters. Gaps exist for Southern 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).

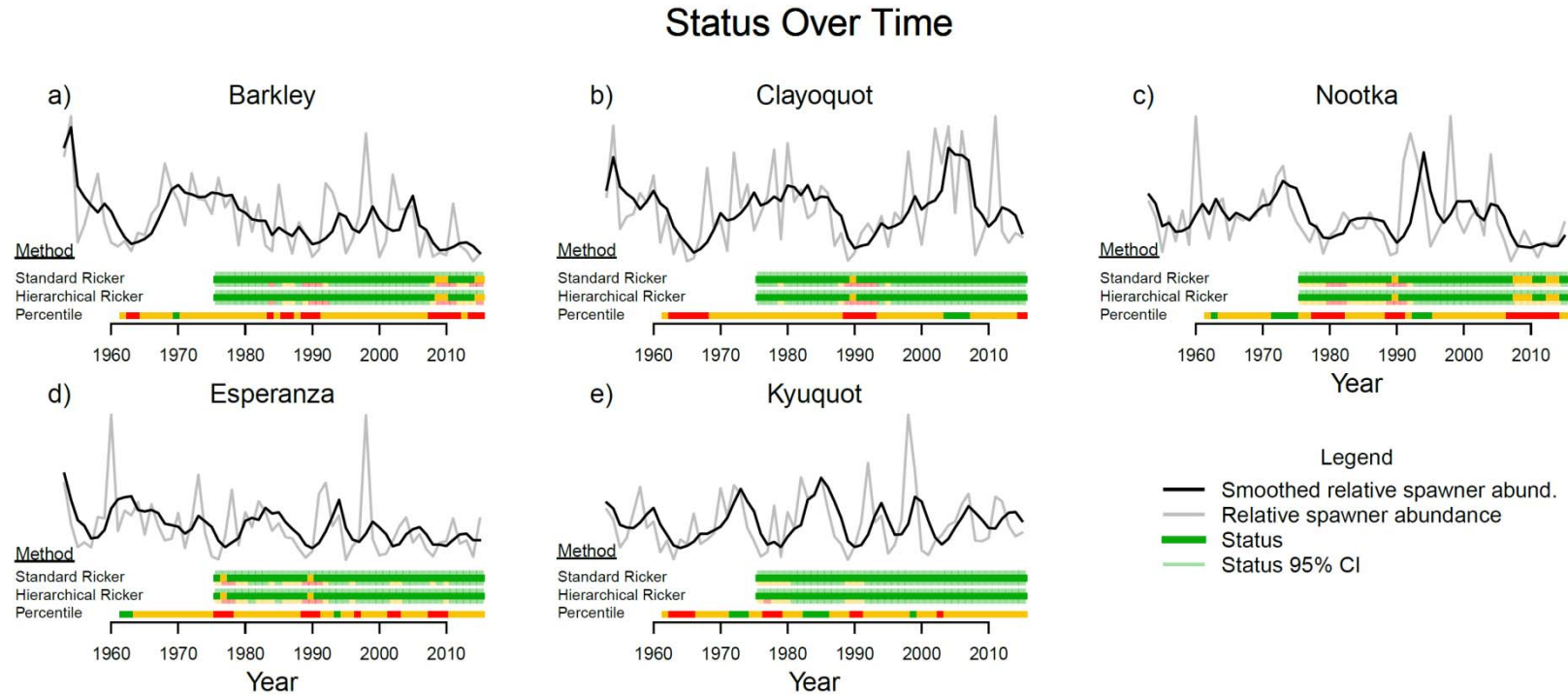


Fig. 10. Standardized raw and generational average escapements across CUs, with status as measured by percentile benchmarks indicated by horizontal coloured bars, and status as measured by standard Ricker benchmarks model indicated by vertical coloured bars. The coloured proportions of each vertical bar represent the probability that status falls within each zone. The location of the amber zone is constant over years resulting in shift in location of the bar upwards when there is a high probability of green status, and downwards when there is a high probability of red status.

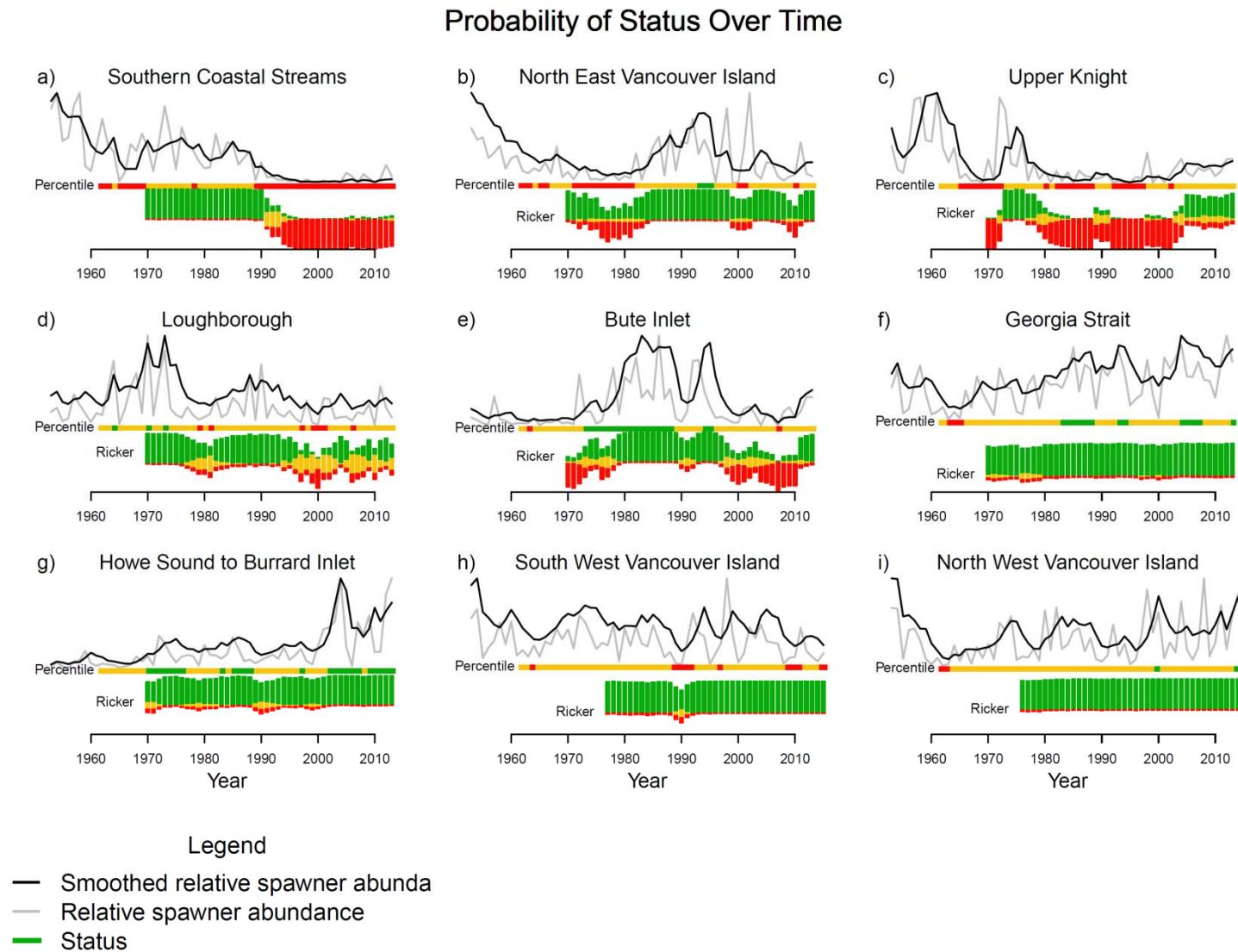


Fig. 11. Standardized raw and generational average escapements across SMUs, with status as measured by percentile benchmarks indicated by horizontal coloured bars, and status as measured by standard Ricker benchmarks model indicated by vertical coloured bars. The coloured proportions of each vertical bar represent the probability that status falls within each zone. The location of the amber zone is constant over years resulting in shift in location of the bar upwards when there is a high probability of green status, and downwards when there is a high probability of red status.

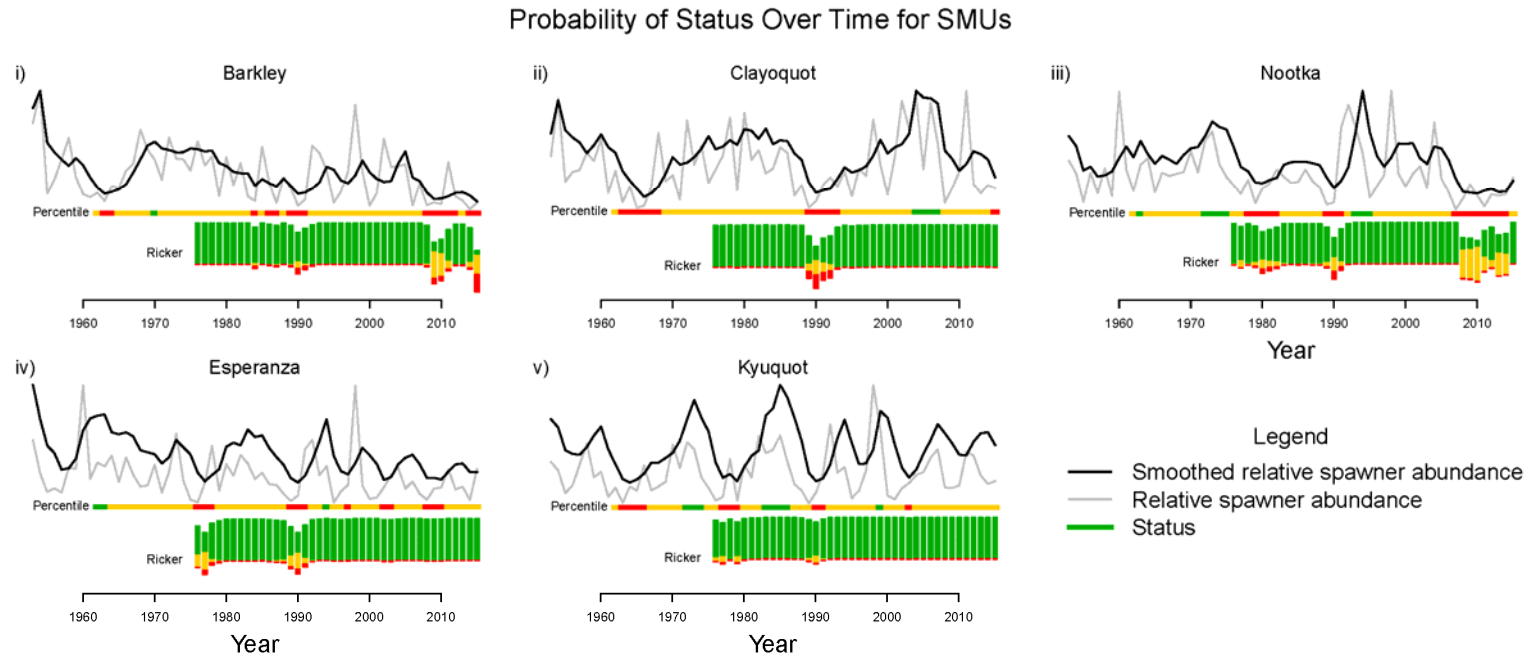


Fig. 12. Estimated Ricker  $\alpha$  values for CUs using a recursive Bayes model, which allows  $\alpha$  to vary over time within a given CU. Grey shaded polygons indicate 95% credible intervals based on posterior distribution of estimated  $\alpha$  values.

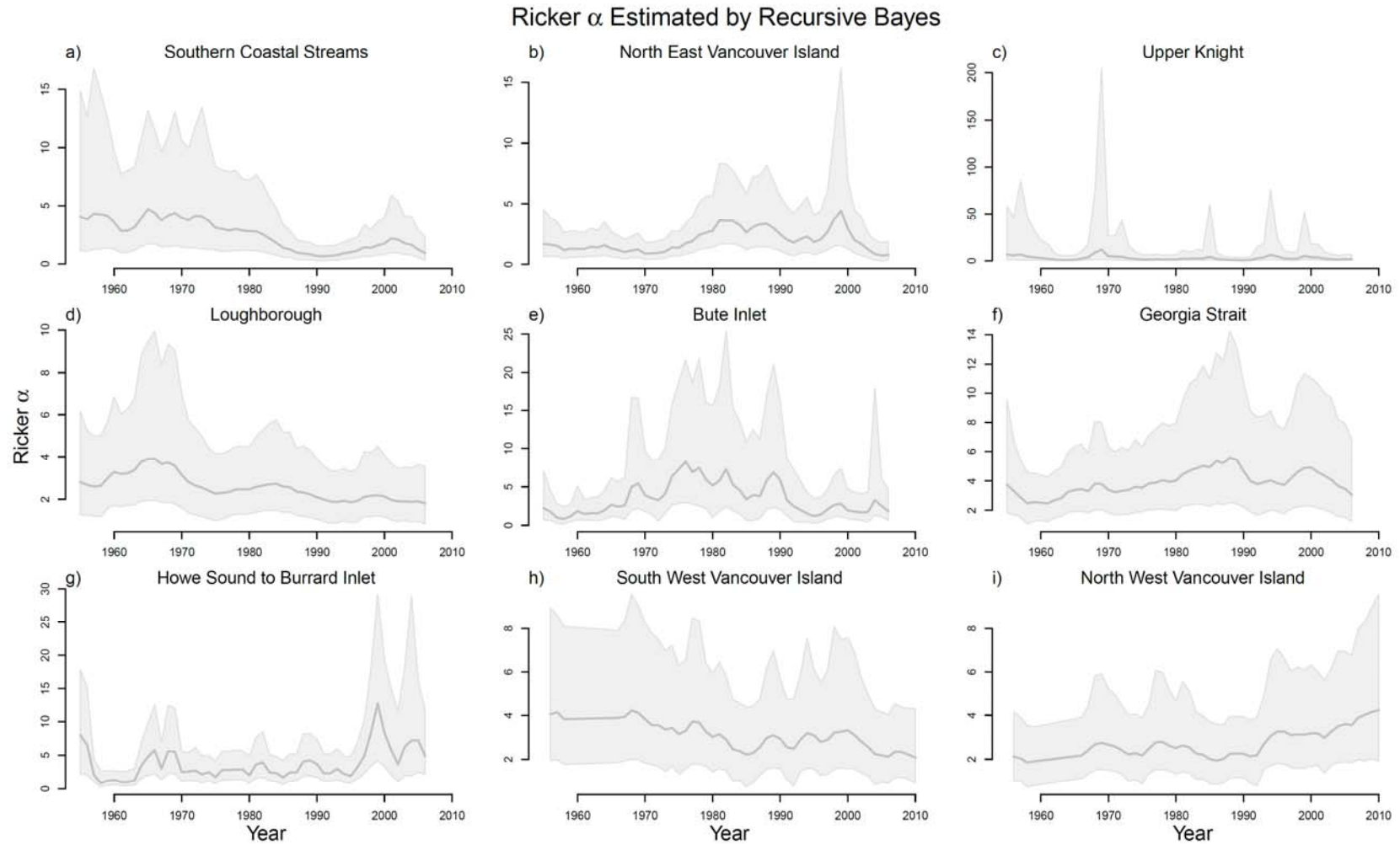


Fig. 13. Estimated Ricker  $\alpha$  values for SMUs using a recursive Bayes model, which allows  $\alpha$  to vary over time within a given CU. Grey shaded polygons indicate 95% credible intervals based on posterior distribution of estimated  $\alpha$  values.

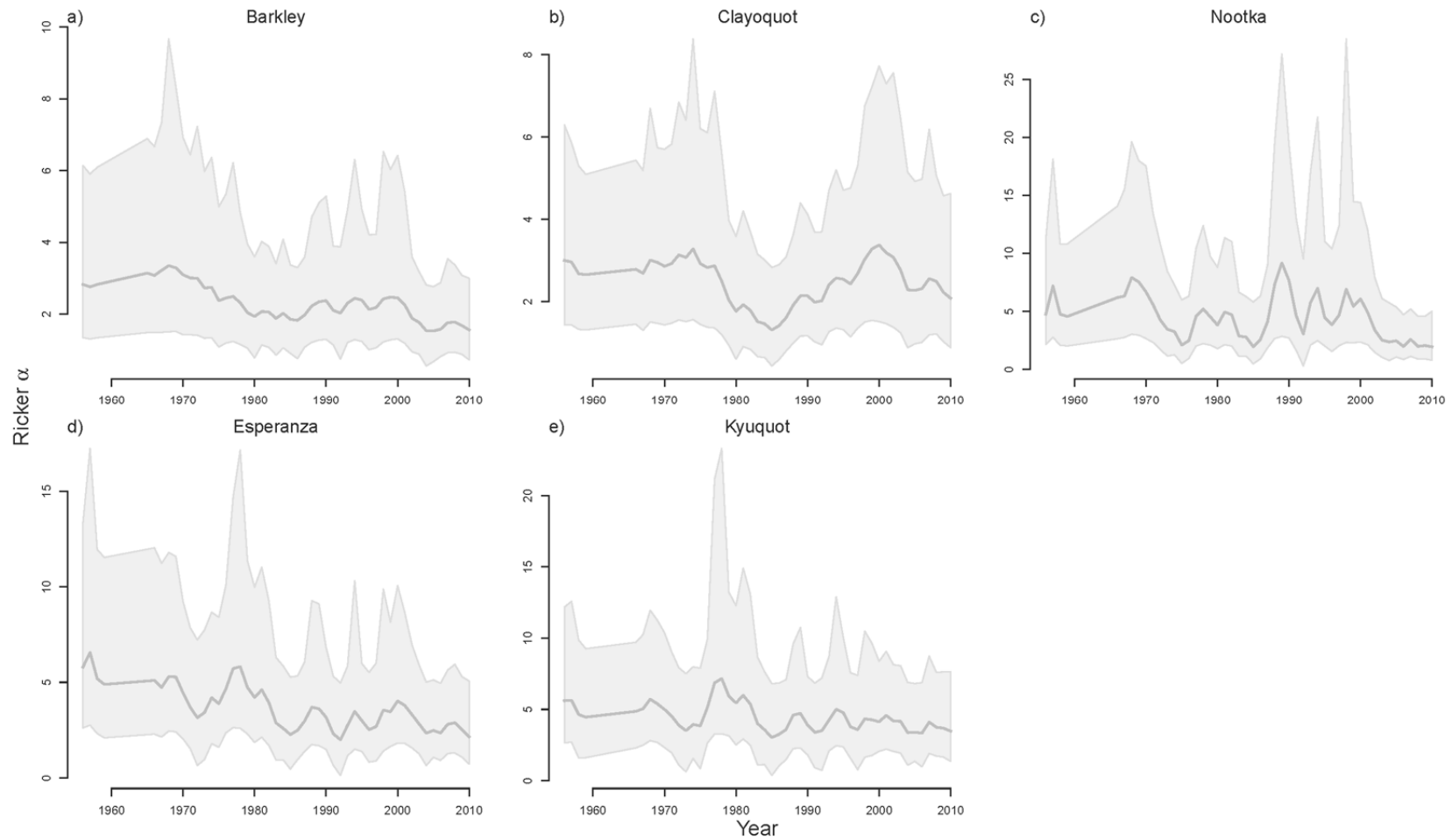


Fig. 14. Schematic of simulation model used to evaluate benchmark performance.

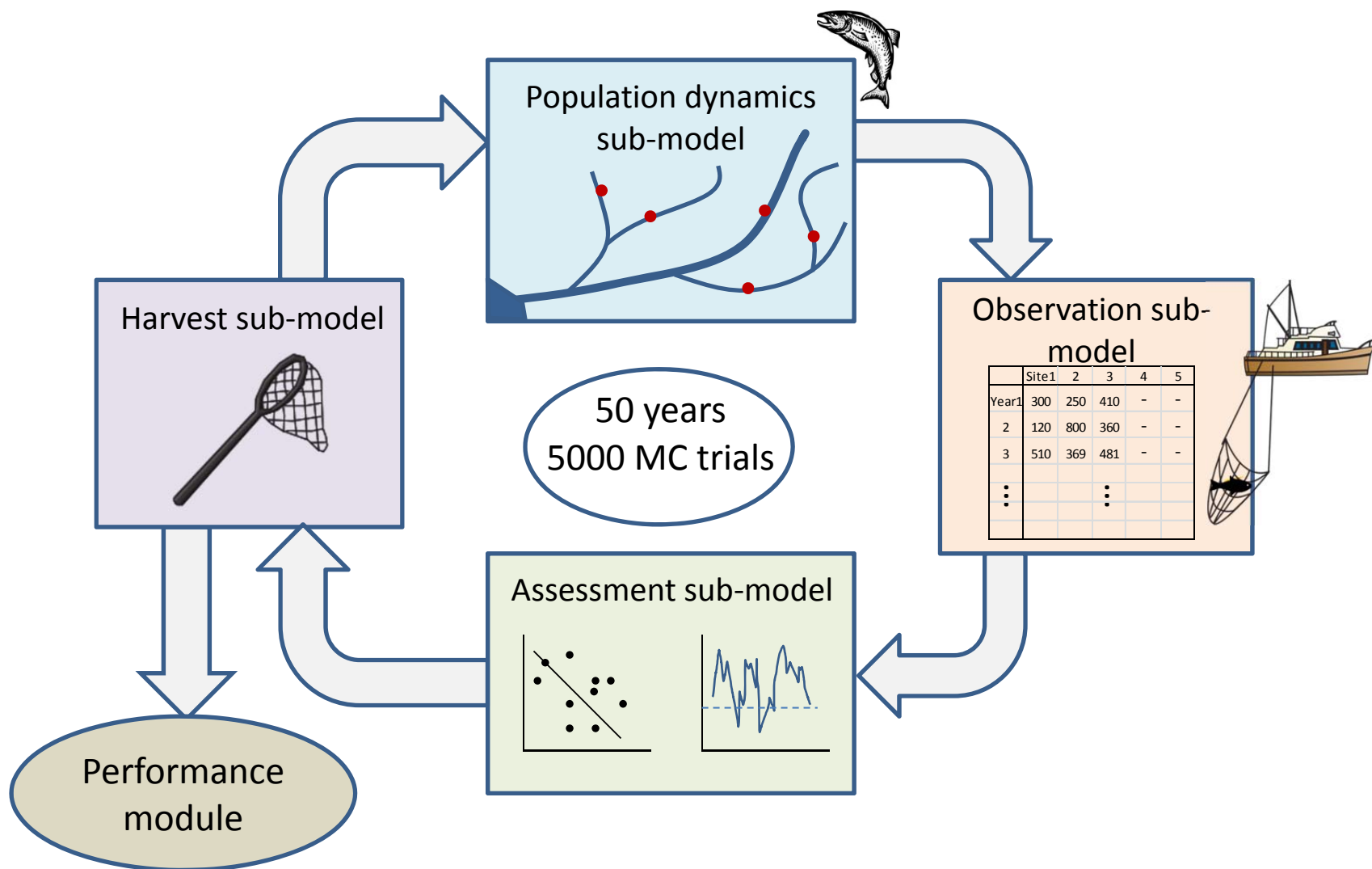


Fig. 15 Time-series of observed spawner abundances (black line) and “true” spawner abundances (grey line) and benchmarks for one Monte Carlo trial of an example CU. Benchmarks are estimated annually base on all data up until that year: annual estimates of 80%  $S_{MSY}$  (upper benchmark, green dashed 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 the mean percent error from the “true” 80%  $S_{MSY}$ .

BI

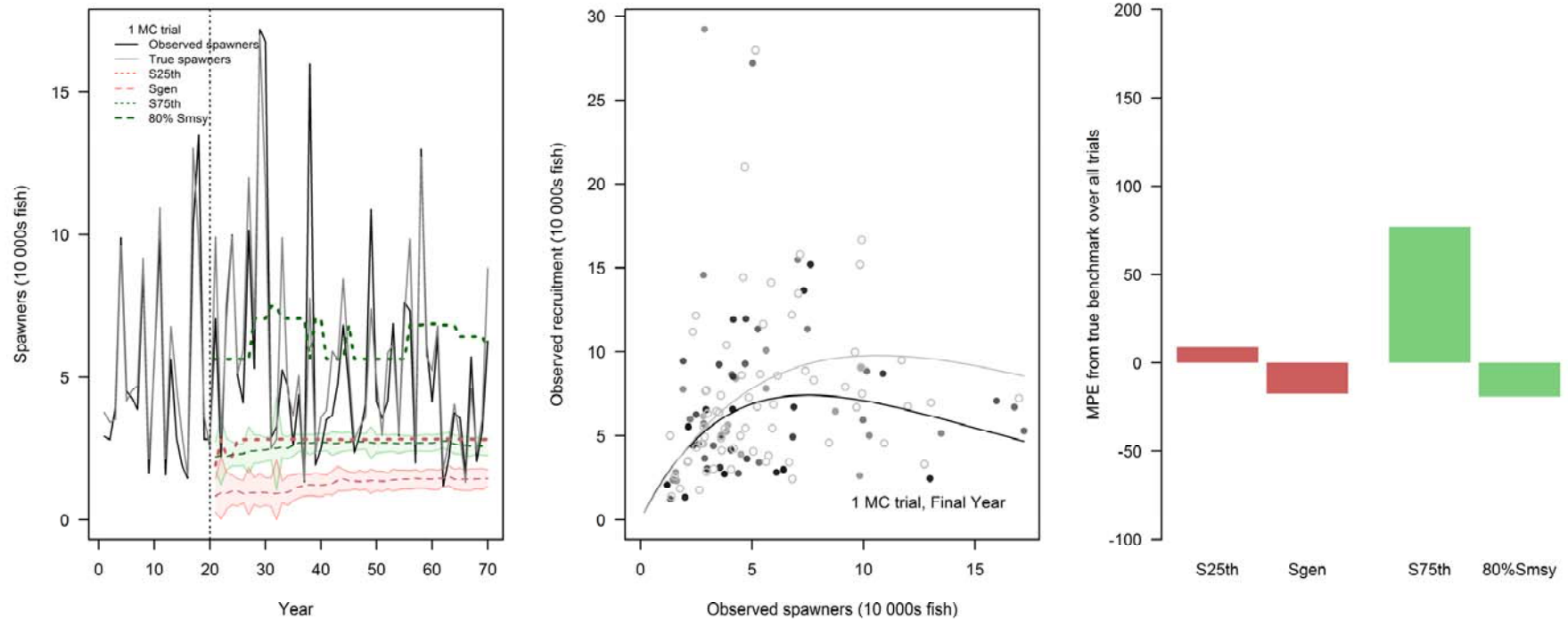


Fig. 16. Difference in the mean percent error, MPE, of estimated lower benchmark ( $S_{\text{gen}}$ , (a), and  $S_{25\text{th}}$  (b)) and the “true” lower benchmarks (“true”  $S_{\text{gen}}$ ), between sensitivity analyses listed on the  $x$ -axis and the base case scenario. Black bars are analyses 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: 132% (a) and 586% (b).

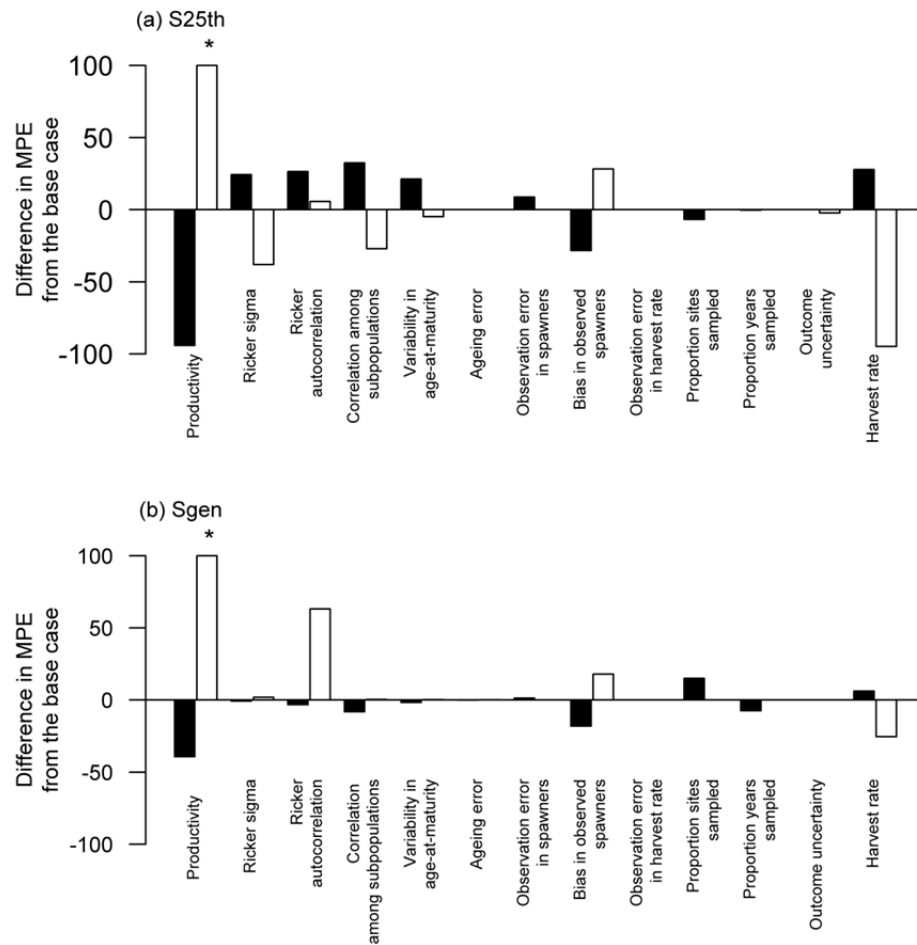


Fig. 17. Sensitivity indices of the effects of individual variables (mean elemental effect, x-axis) and interactions among variables (standard deviation in elemental effects, y-axis). Indices were derived from the Morris method, a global sensitivity analyses for the mean percent error of estimated lower benchmarks ( $S_{\text{gen}}$  (a), and  $S_{25\text{th}}$  (b)) from “true” benchmarks. Input variables with values  $>100$  on either axis are labelled.

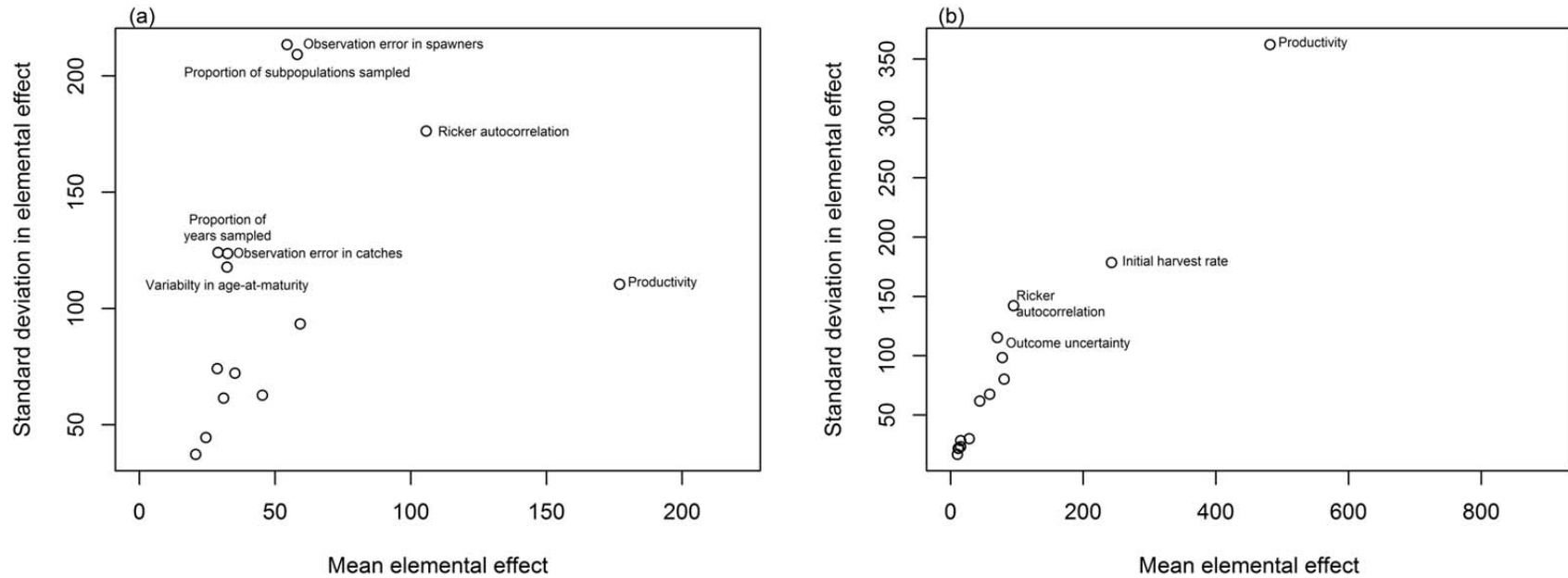


Fig. 18. Mean percent error, MPE, of the estimated lower benchmark (either  $S_{25th}$  (a), and  $S_{gen}$  (b)) from the “true” lower benchmark,  $S_{gen}$  along a gradient in harvest rates ( $x$ -axis) and productivities ( $y$ -axis) derived from a simulation model of a hypothetical Chum Salmon CU. Symbols indicate MPE of Inner South Coast 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.

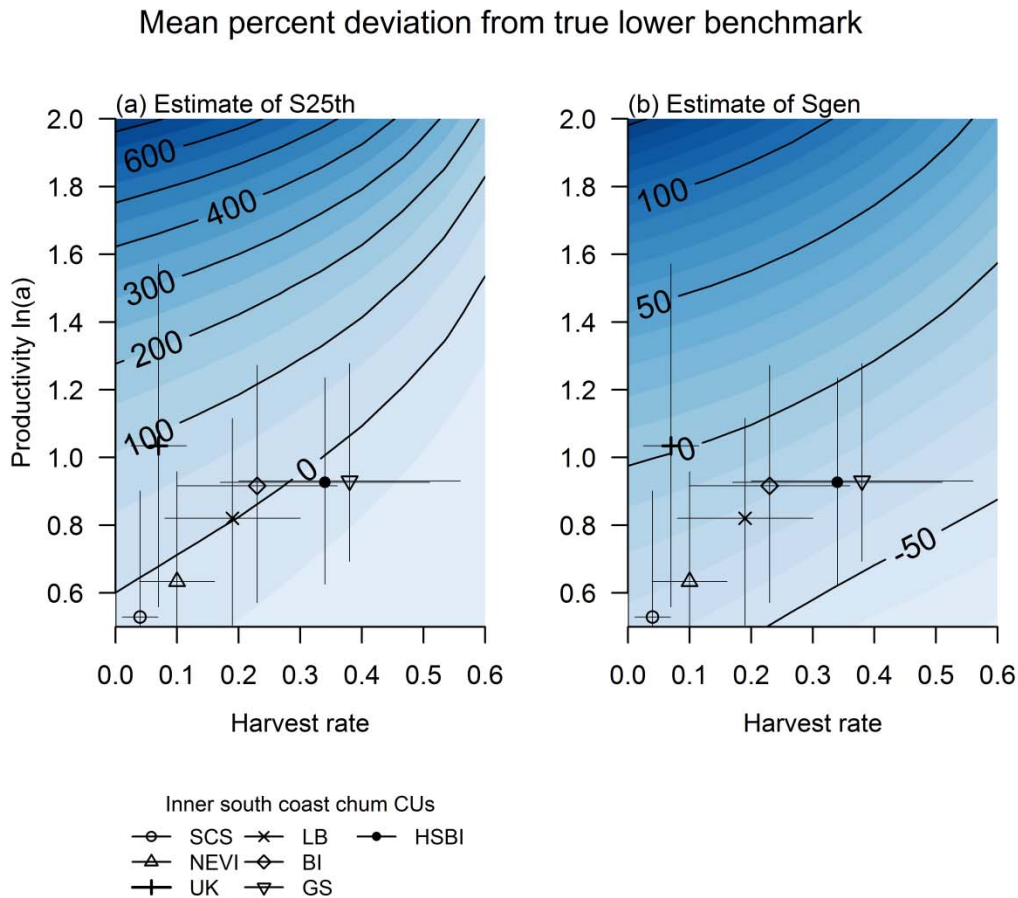


Fig. 19. Mean percent error, MPE, of the estimated lower benchmark based on the 50<sup>th</sup> percentile of observed abundances,  $S_{50th}$ , (a), and  $S_{gen}$  (b), from the “true” lower benchmark,  $S_{gen}$  along a gradient in harvest rates ( $x$ -axis) and productivities ( $y$ -axis), derived from a simulation model of a hypothetical Chum Salmon CU. See the caption for Fig. 18 for an explanation of symbols, lines, and abbreviations.

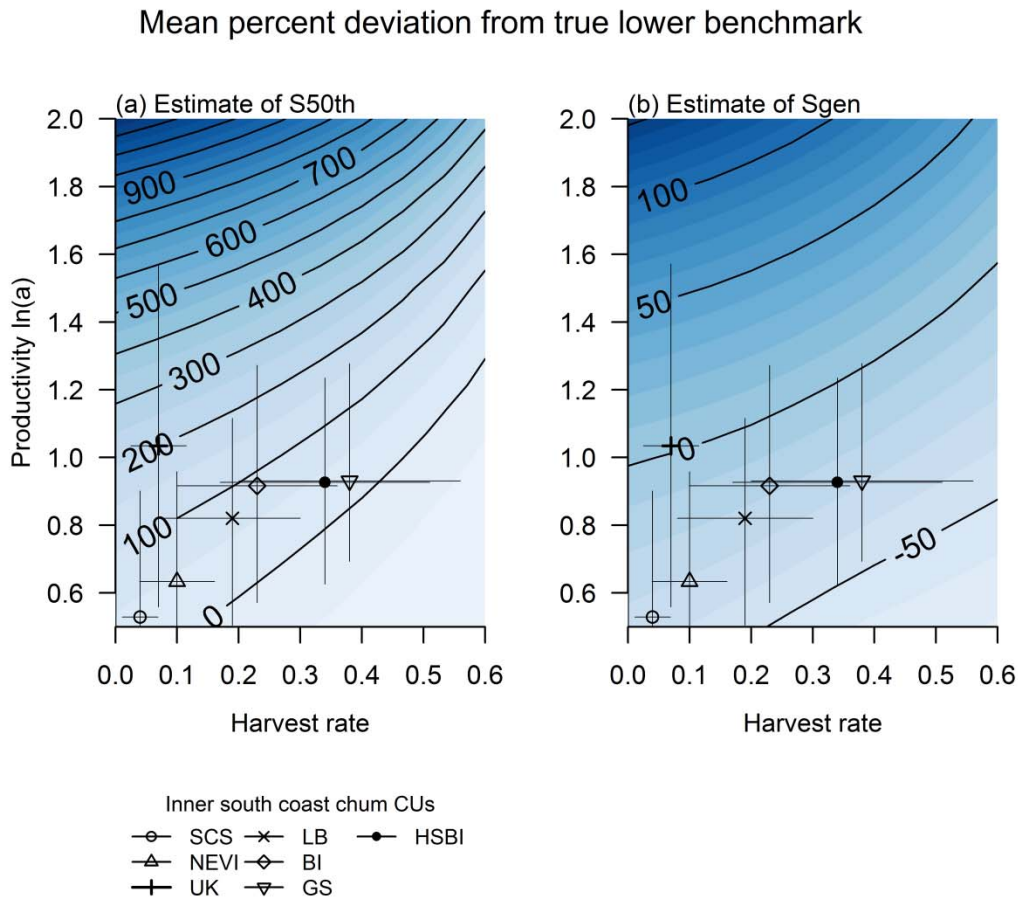


Fig. 20. Mean percent error, MPE, of the estimated lower benchmark ( $S_{25th}$  (a), and  $S_{gen}$  (b)) from the “true”  $S_{gen}$  lower benchmark along a gradient in harvest rates ( $x$ -axis) and productivities ( $y$ -axis) derived from a simulation model of a hypothetical Chum Salmon CU. Symbols indicate MPE of West Coast of Vancouver Island CUs and SMUs assuming productivities estimated from hierarchical Ricker models and mean harvest rates over available time-series for each CU and SMU. Y-error bars represent the 95% credible intervals of the estimate of productivity. X-error bars are the standard deviation of historical harvest rates. SWVI is the Southwest Vancouver Island CU and NWVI is the Northwest Vancouver Island CU, which are indicated with dark black error bars. The remaining stocks are SMUs, indicated with grey error bars.

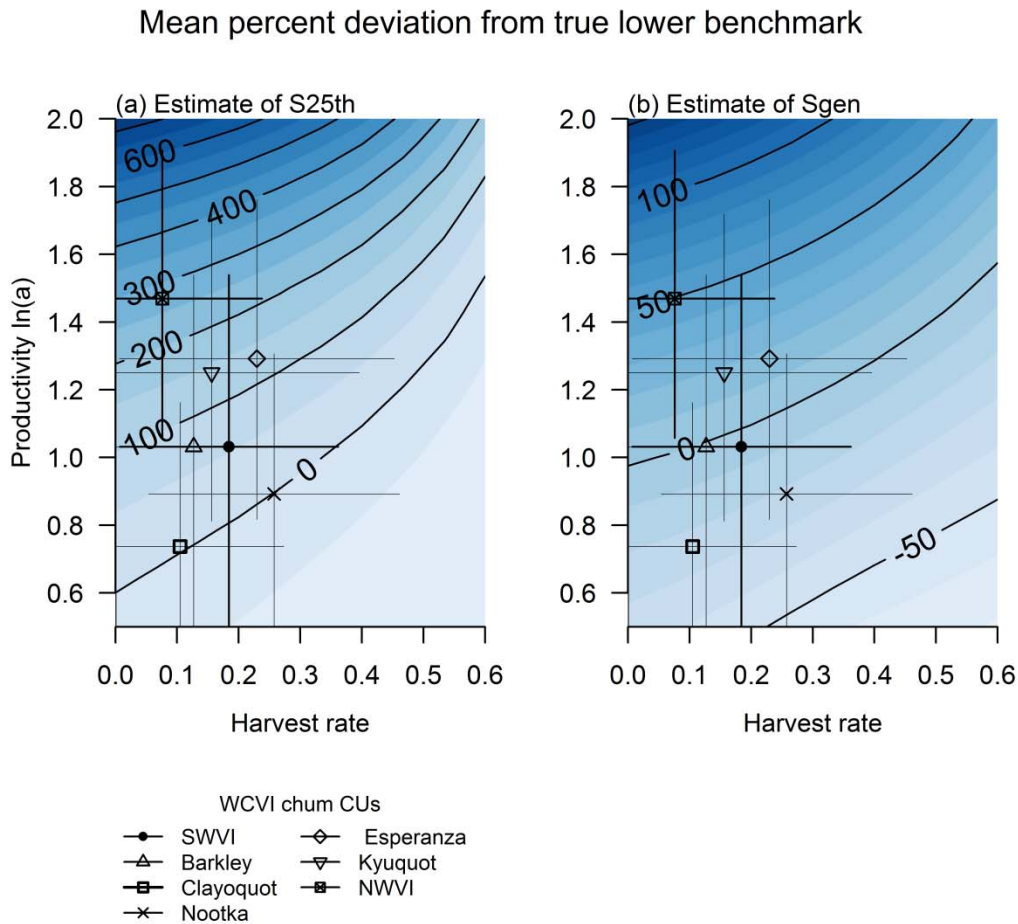


Fig. 21. Mean percent error, MPE, of the estimated upper benchmark ( $S_{75th}$  (a), and 80%  $S_{MSY}$  (b)) from the “true” upper benchmark, 80%  $S_{MSY}$  value along a gradient in harvest rates ( $x$ -axis) and productivities ( $y$ -axis) derived from a simulation model of a hypothetical Chum Salmon CU. Symbols indicate MPE of Inner South Coast 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.

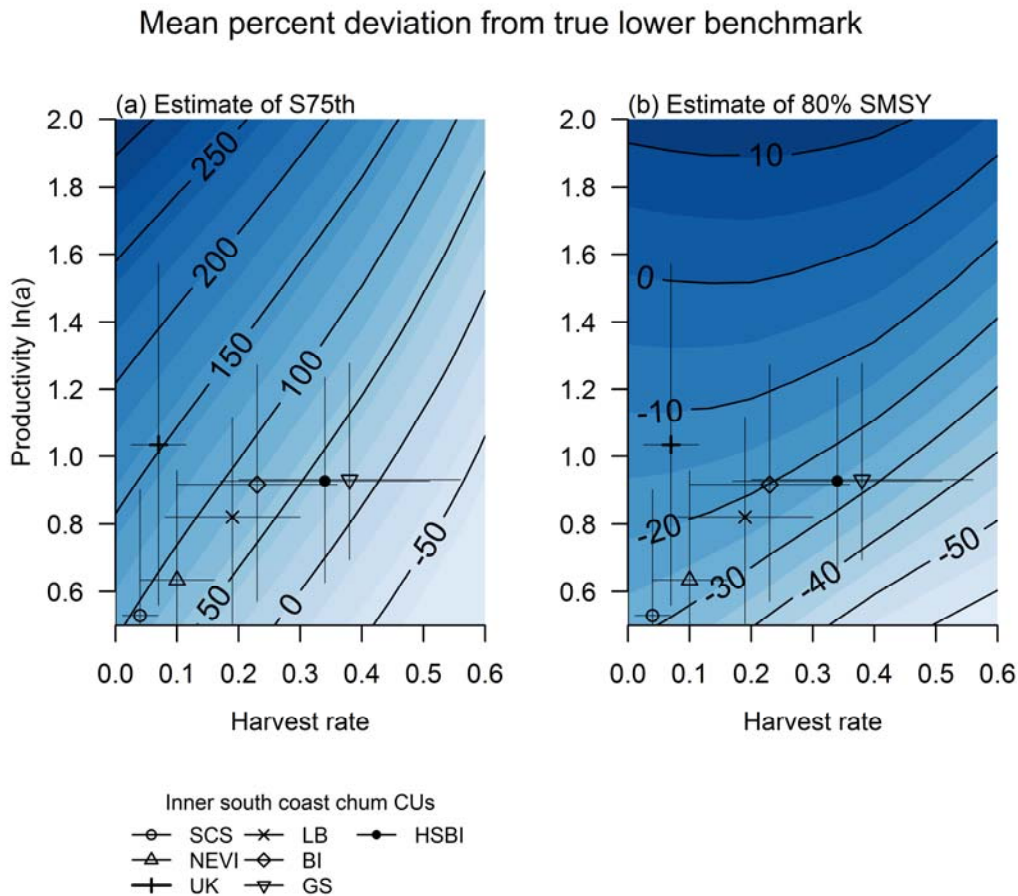


Fig. 22. Mean percent error, MPE, of the estimated upper benchmark ( $S_{75th}$  (a), and 80%  $S_{MSY}$  (b)) from the “true” upper benchmark, 80%  $S_{MSY}$  value along a gradient in harvest rates ( $x$ -axis) and productivities ( $y$ -axis) derived from a simulation model of a hypothetical Chum Salmon CU. Symbols indicate MPE of West Coast of Vancouver Island CUs and SMUs 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. SWVI is the Southwest Vancouver Island CU and NWVI is the Northwest Vancouver Island CU, which are indicated with dark error bars. The remaining stocks are SMUs, indicated with grey error bars.

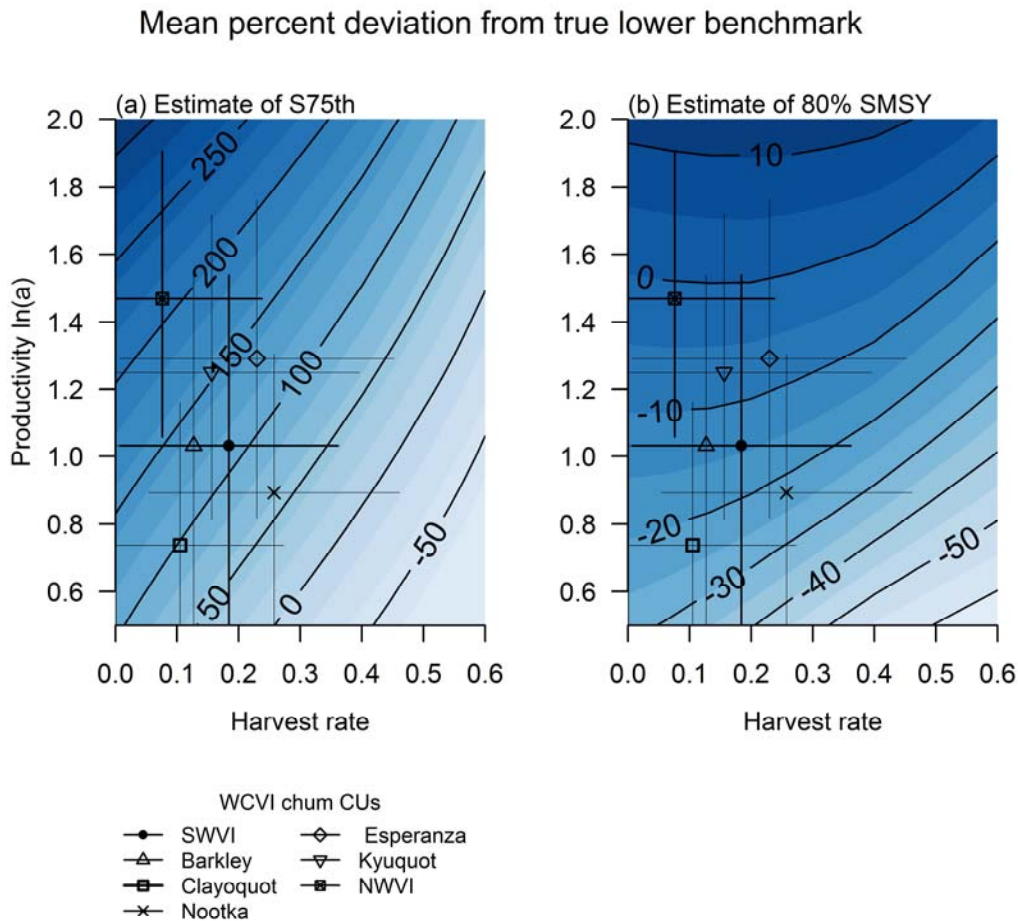


Fig. 23. Mean raw error of estimated Ricker  $\log_e(\alpha)$  values (left panel, labeled MRE in Ricker a) and carrying capacity values (right panel, labelled MRE in Ricker b) along gradients in true  $\log_e(\alpha)$  values (productivity) and harvest rates derived from a simulation model of a hypothetical Chum Salmon CU.

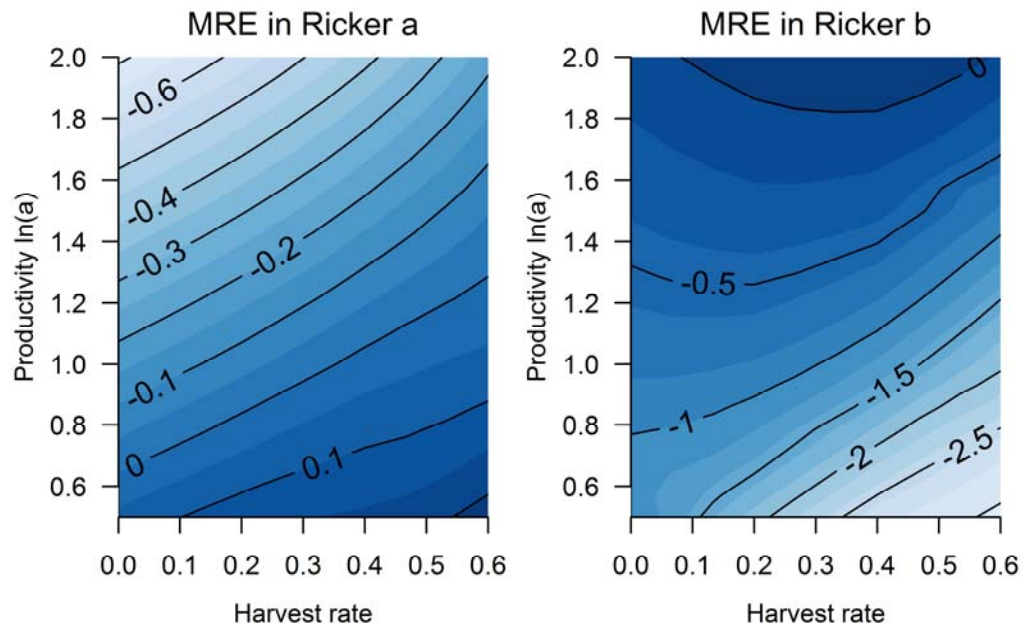


Fig. 24. Contrast in spawner escapement data (maximum escapement/minimum escapement) over gradients in true  $\log_e(\alpha)$  values (productivity) and harvest rates derived from a simulation model of a hypothetical Chum Salmon CU.

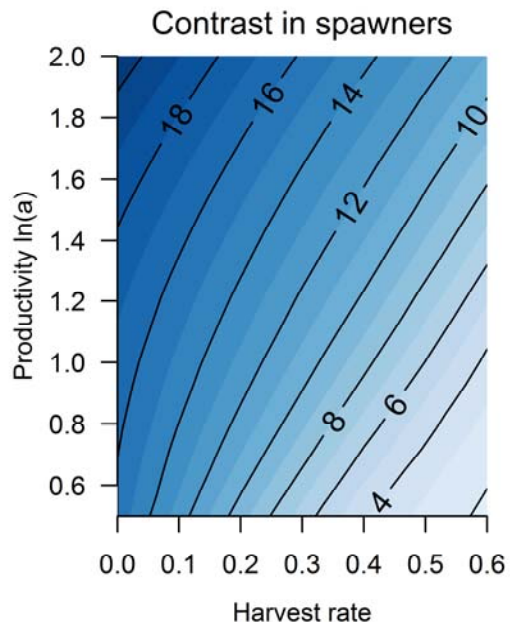


Fig. 25. Mean percent error, MPE, of the estimated lower benchmark ( $S_{25th}$  (a), and  $S_{gen}$  (b)) from the “true”  $S_{gen}$  value along a gradient in bias in estimated spawner abundance ( $x$ -axis) and observation errors in spawner abundances (SD,  $y$ -axis) derived from a simulation model of a hypothetical Chum Salmon CU.

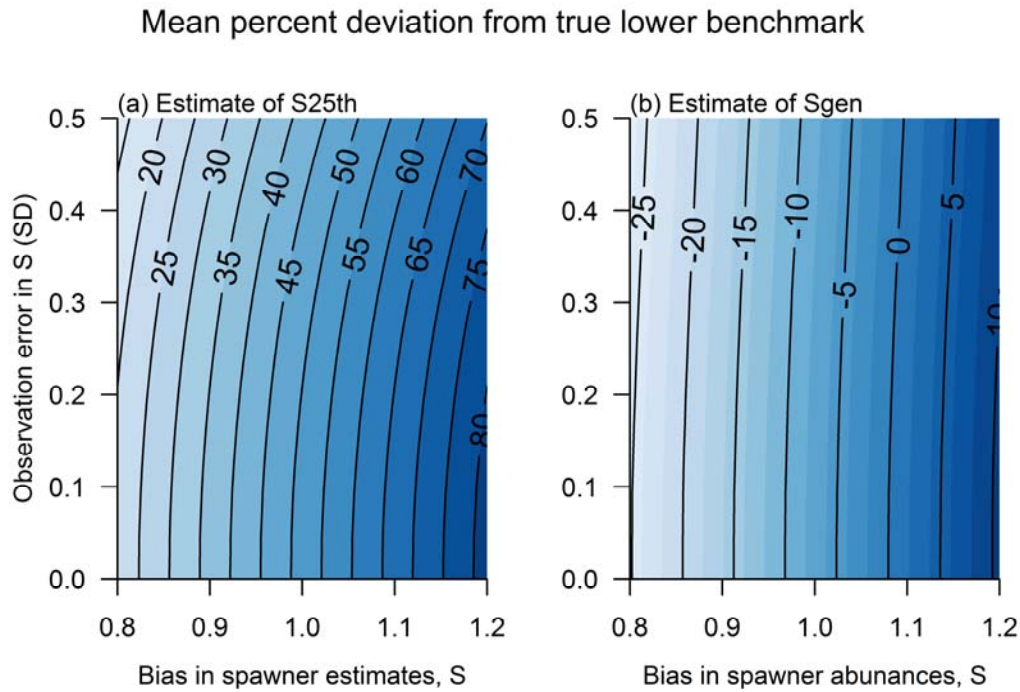


Fig. 26. “True” values of Ricker parameters and benchmarks under assumption of constant spawners at equilibrium,  $S_{eq}$ , (a-e) or constant spawner abundances at maximum recruitment  $S_{max}$  (f-j), with abrupt changes in productivity in year 35 of simulation.

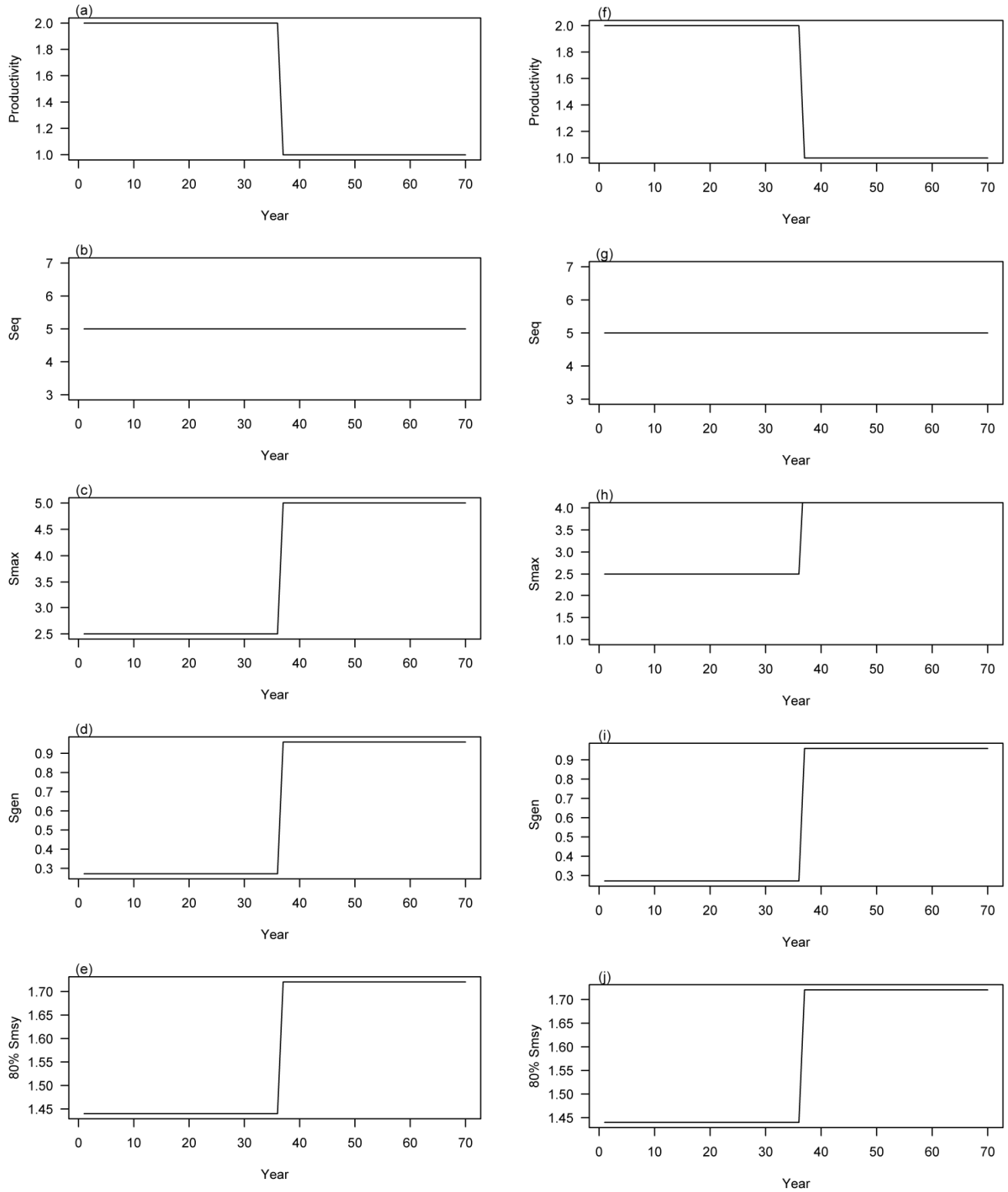


Fig. 27. Box plots of parameter and benchmark values in final year of the simulation averaged over all Monte Carlo trials, using either the first 30 years of data (dark grey boxes), all 70 years of data (light grey boxes), or the most recent 30 years of data (white boxes). Boxes represent the lower quartiles, 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.

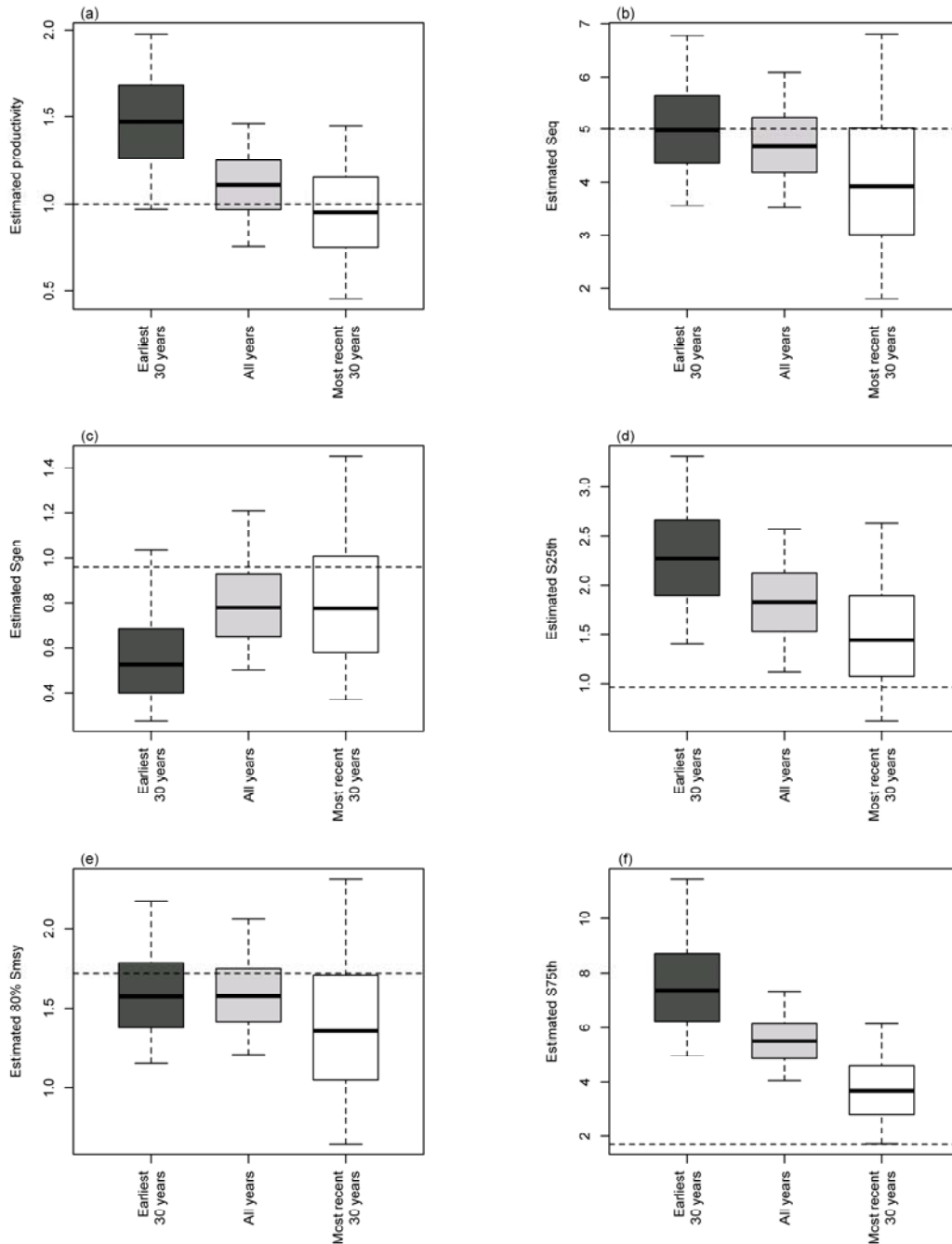


Fig. 28. Mean percent error of estimated Ricker parameters and lower and upper benchmarks from the “true” values, using only the first 30 years of data (black bars), all 70 years (grey bars), and the most recent 30 years (white bars) under the assumption that  $S_{eq}$  remained constant over time (a), or  $S_{max}$  remained constant (b).

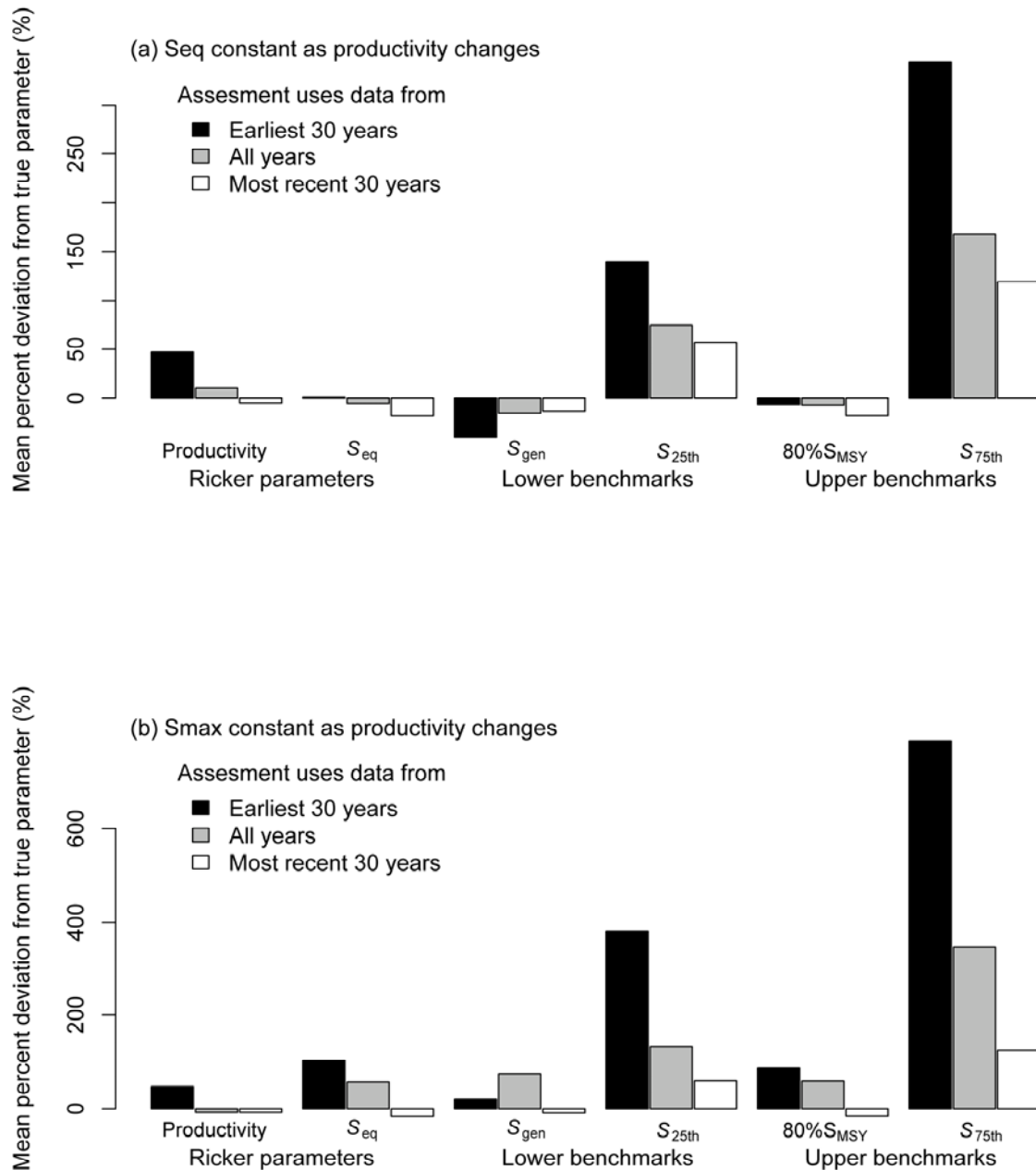


Fig. 29. Caption as for Fig. 27, except  $S_{\max}$  was held constant and  $S_{\text{eq}}$  declined in simulations as productivity varied.

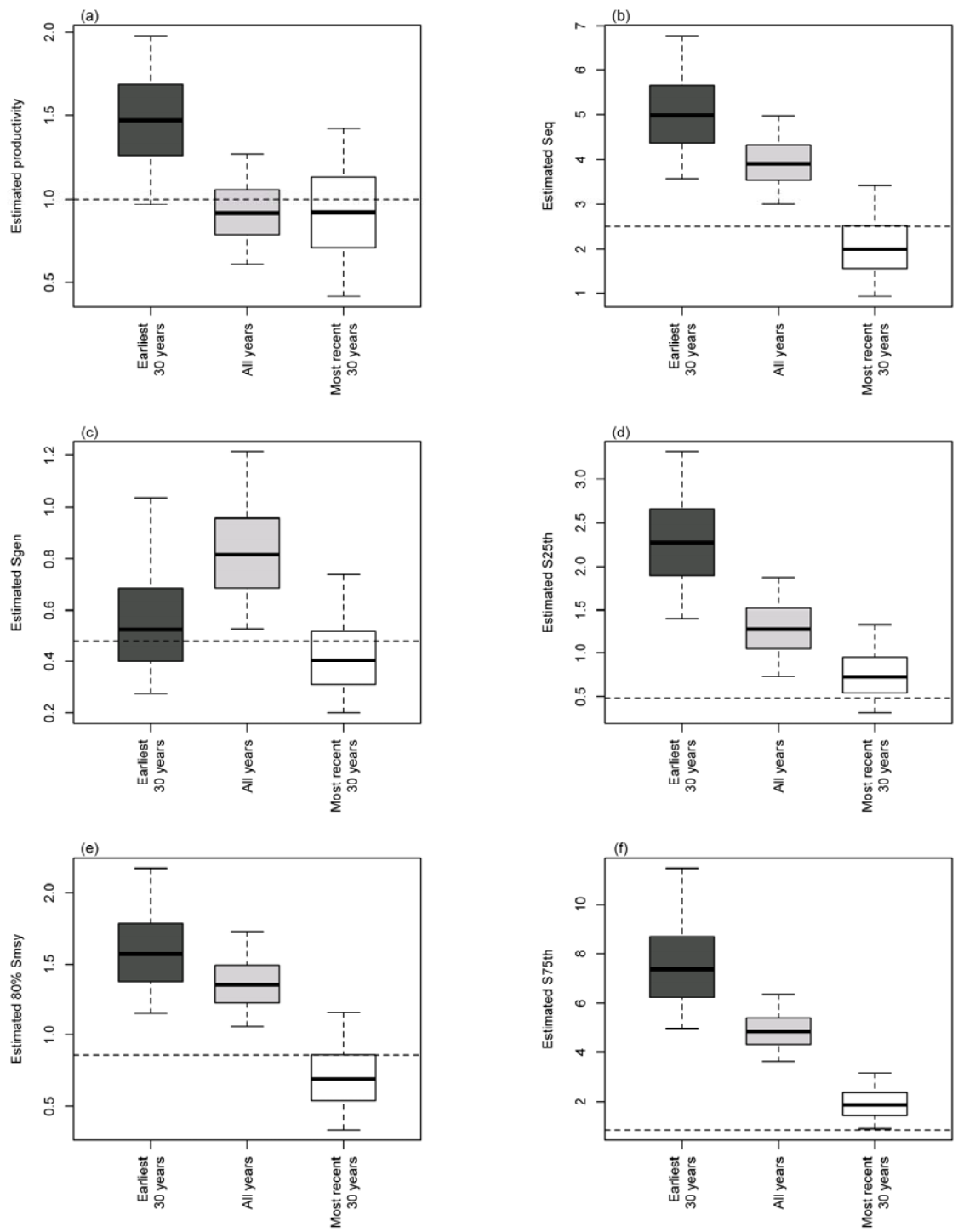
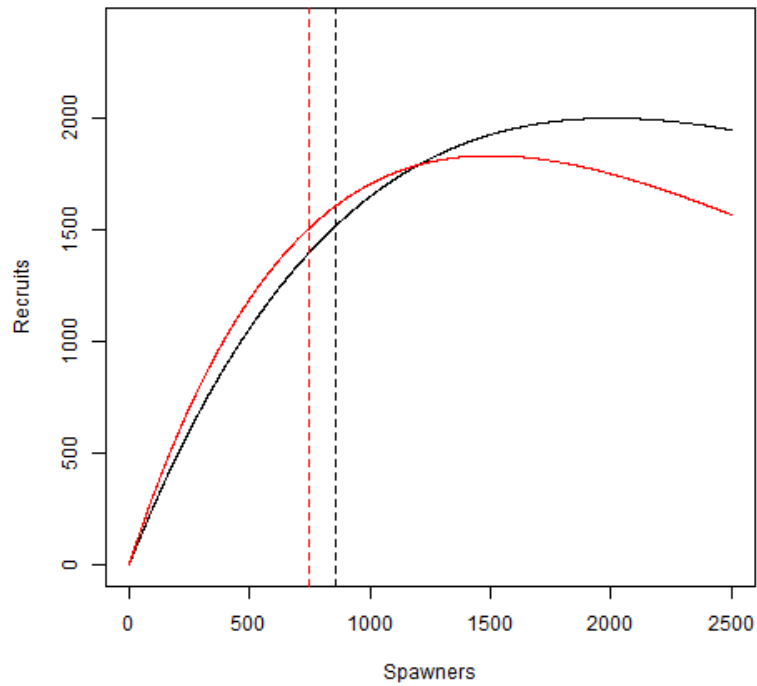


Fig.30. Stock-recruitment curves for a hypothetical CU under a base case of moderate productivity (black curve), and scenario where productivity is over-estimated and carrying capacity is underestimated, as occurs for time-series biases (red curve). Dashed lines represent  $S_{\text{gen}}$  benchmarks for the base case (black) and biased parameter estimates (red).



## APPENDICES

### Appendix A

Fig. A1. Priors and posteriors for Ricker  $\alpha$  parameters for both the basic and hierarchical Ricker model, for CUs.

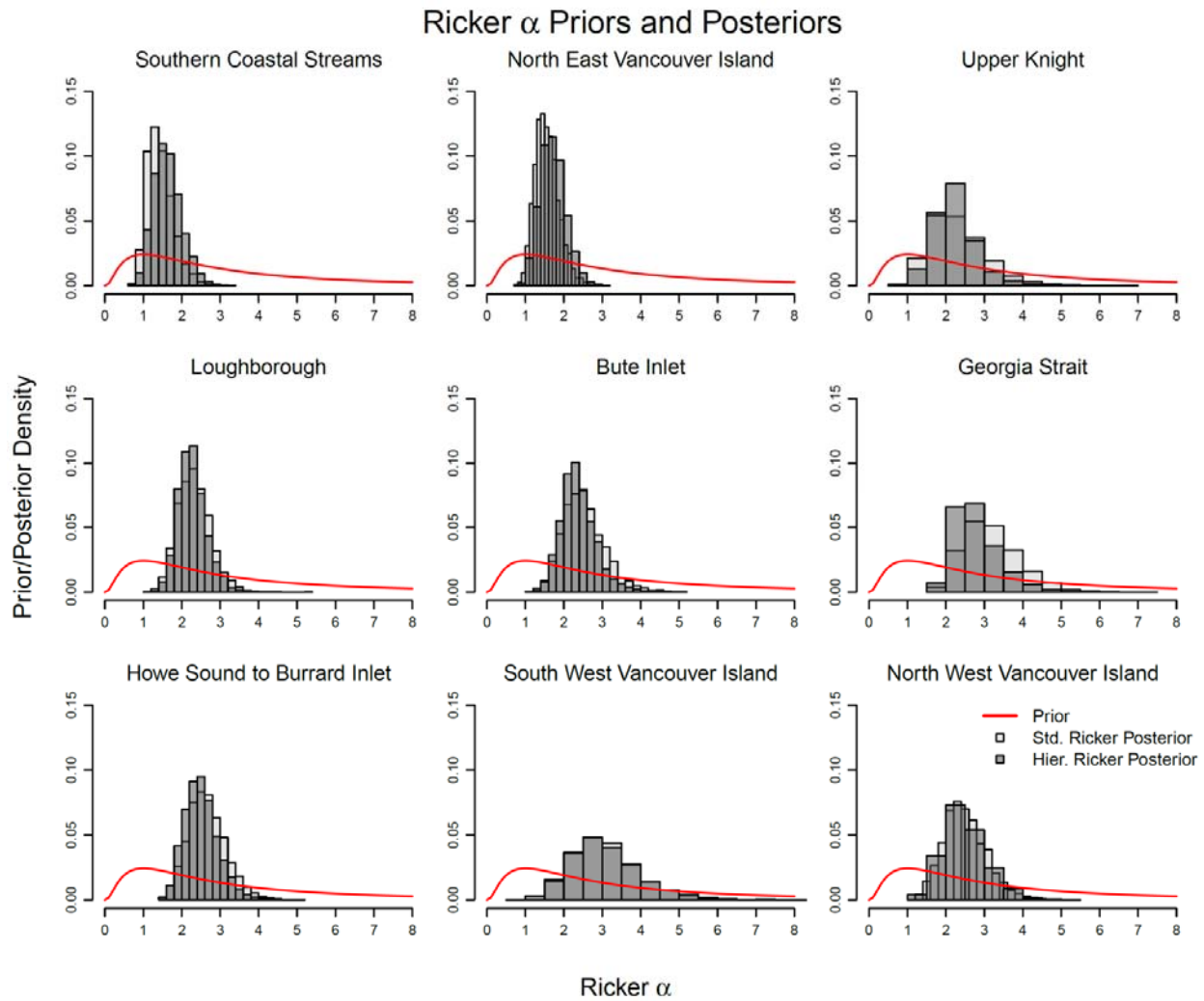
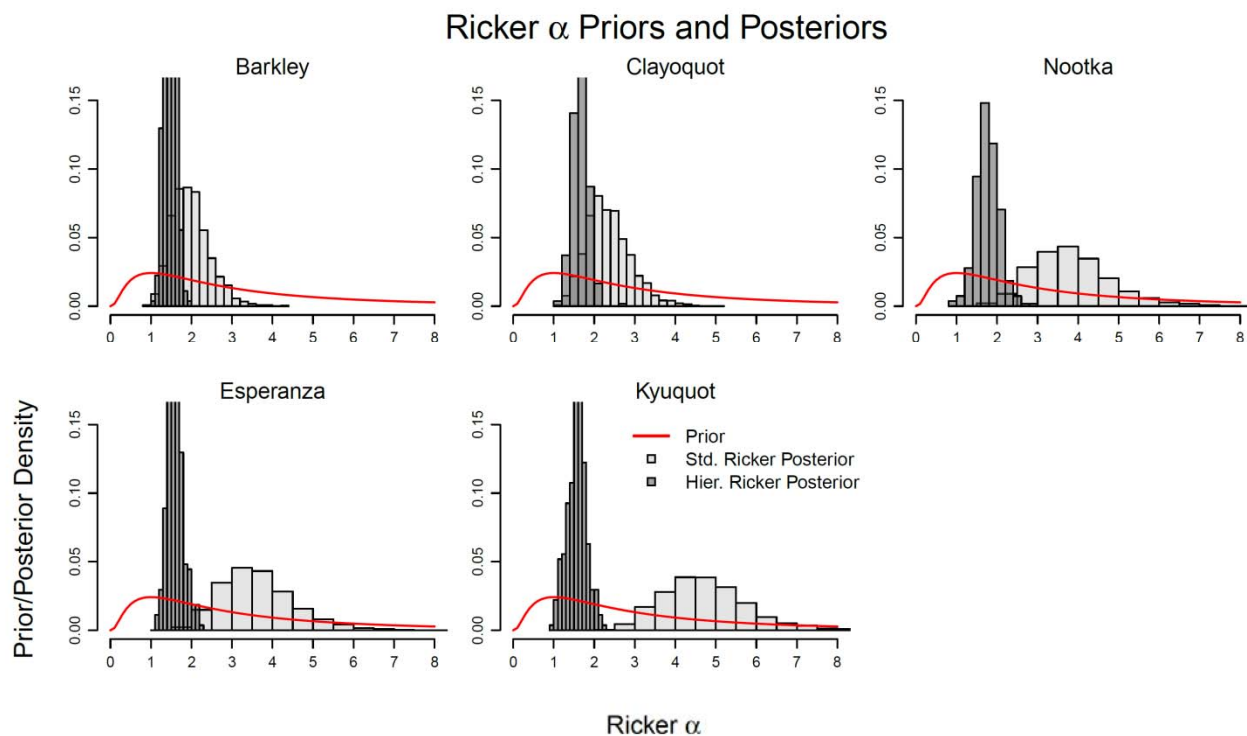


Fig. A2. Priors and posteriors for Ricker  $\alpha$  parameters for both the basic and hierarchical Ricker model, for Southwest Coast Vancouver Island SMU's



## Appendix B

Two prior formulations on Ricker  $\beta$ , via its reciprocal:  $S_{max}$ , were used, as described in equations 4a and 4b:

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

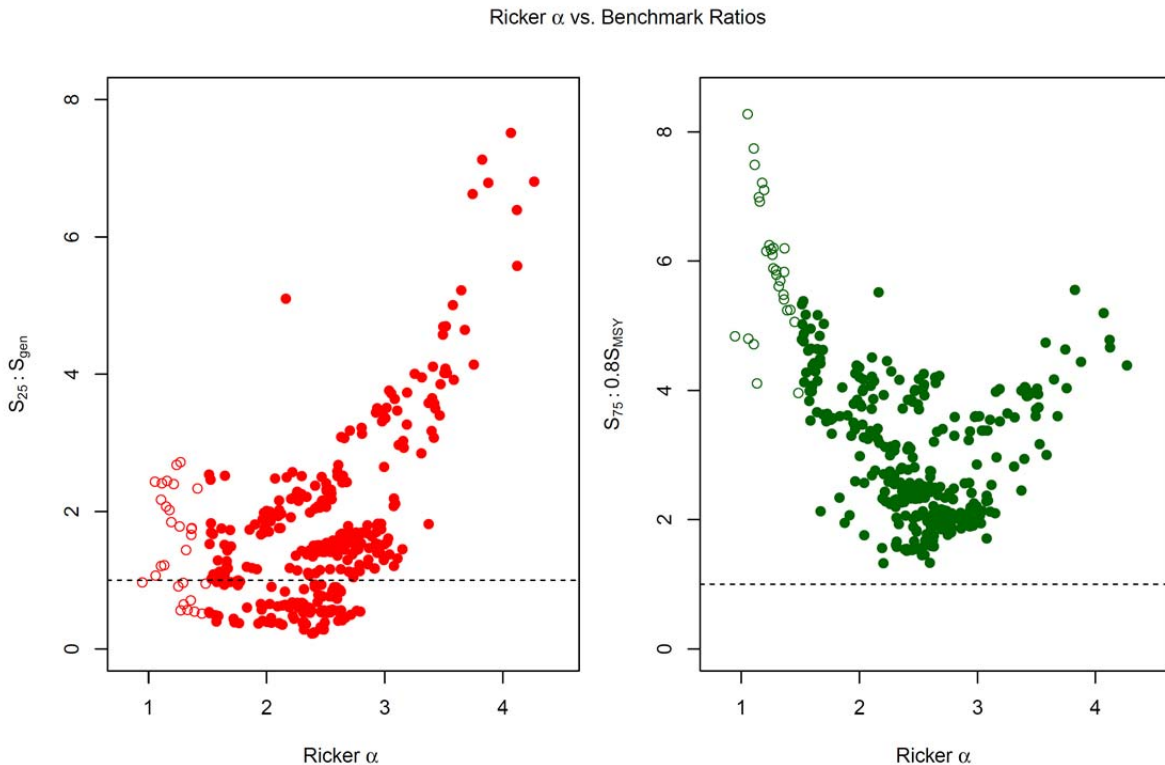
$$(4b) S_{max} \sim \text{lognormal}(\log(\text{mean}(S_{obs})), \tau_S), \tau_S = 1/\log(CV^2 + 1)$$

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 Chum Salmon CUs in southern BC, 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. A comparison of estimates using either prior can be seen in Fig 2.

## Appendix C

We explored an alternative way to compare Ricker-based and percentile-based benchmarks by comparing the ratio of the percentile-based benchmark to the Ricker benchmark. If this number is above one (dashed line in Fig. C1) the percentile benchmark is higher (and therefore more precautionary) than the Ricker-based benchmark. These figures show that the upper percentile benchmark is always higher than the Ricker-based lower benchmark ( $S_{gen}$ ), while at low/moderate productivities, percentile lower benchmarks can dip below  $S_{gen}$ . There is an interesting relationship between the Ricker  $\alpha$  parameter and these ratios. For the lower benchmarks, higher productivity is associated with much more precautionary percentile benchmarks, compared to Ricker based benchmarks. While for upper benchmarks, both very low and very high productivity values are associated with cases where percentile benchmarks tend to be precautionary; percentile and Ricker-based upper benchmarks tend to be most similar at intermediate productivity levels.

Fig. C1. Ricker  $\alpha$  (productivity) parameters vs. ratios of percentile-based benchmarks to Ricker-based benchmarks for ISC and WCVI CUs. Left plot shows ratio for lower benchmarks ( $S_{25} : S_{gen}$ ), right plot shows ratio for upper benchmarks ( $S_{75} : 0.8S_{MSY}$ ). Points lying above the dashed line at ratio=1 identify cases where the percentile benchmark is larger (and therefore more precautionary) than the Ricker-based benchmark. The empty circles indicate points with  $\alpha < 1.5$ , which would not have been used to assess status, based on advice from Holt and Ogden (2013).



## Appendix D

### Adaptations of simulation model and parameterization

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 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 harvest rate (0-60%, varied in sensitivity analyses) was applied instead of a harvest control rule with limit and/or target reference points.
- 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 showed similar patterns and are not shown here.
- 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 D1, 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,  $S_{eq}$  (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 J. et al. 2013), and considered a range of plausible autocorrelation coefficients (0 and 0.9) in sensitivity analyses (Table D1). 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 Josh 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.

Table D1. 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 E

Fig. E1. Mean percent error, MPE, of the alternative lower benchmark, 30<sup>th</sup> percentile of observed spawner time-series,  $S_{30th}$  (a), and  $S_{gen}$  (b) from the “true”  $S_{gen}$  lower benchmark along a gradient in harvest rates ( $x$ -axis) and productivities ( $y$ -axis) derived from a simulation model of a hypothetical Chum Salmon CU. Symbols indicate MPE of Inner South Coast 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.

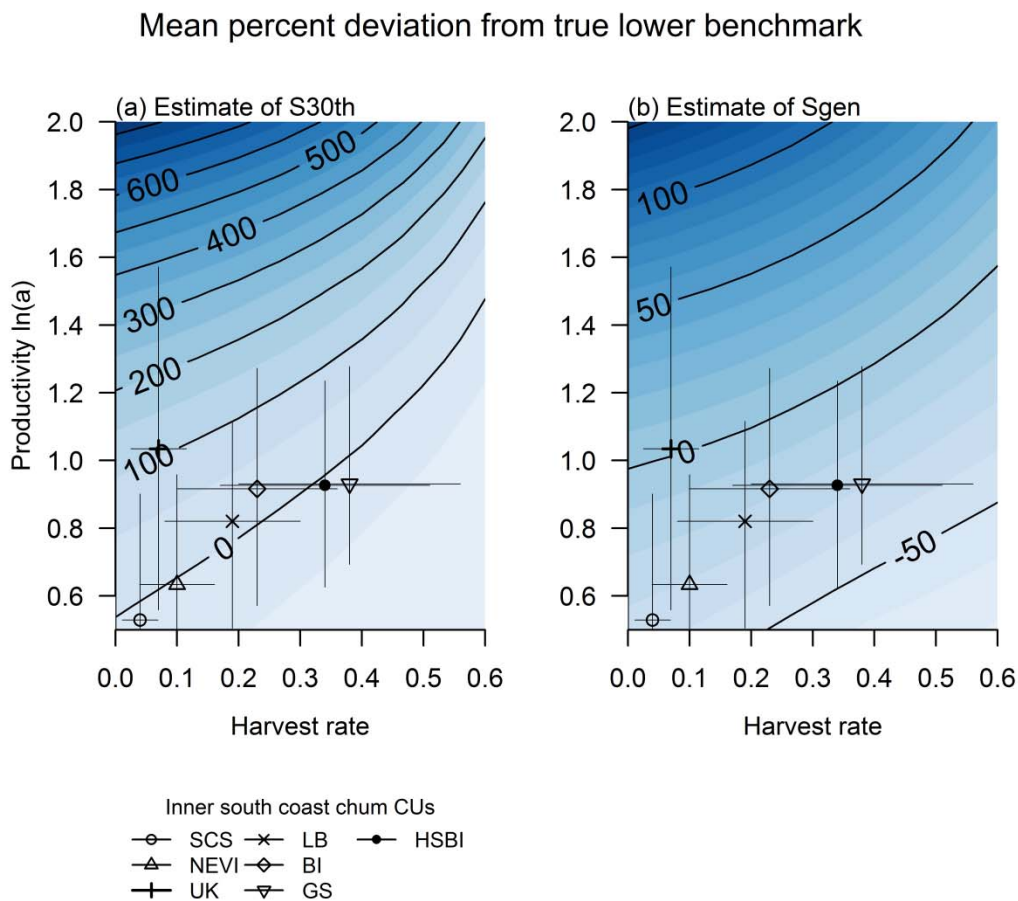


Fig. E2. Same as Fig. E1 except panel (a) depicts performance of the lower benchmark based on the 35<sup>th</sup> percentile of the observed spawner time-series. Panel (b) is the same as in Fig. E1, but is shown here for comparison.

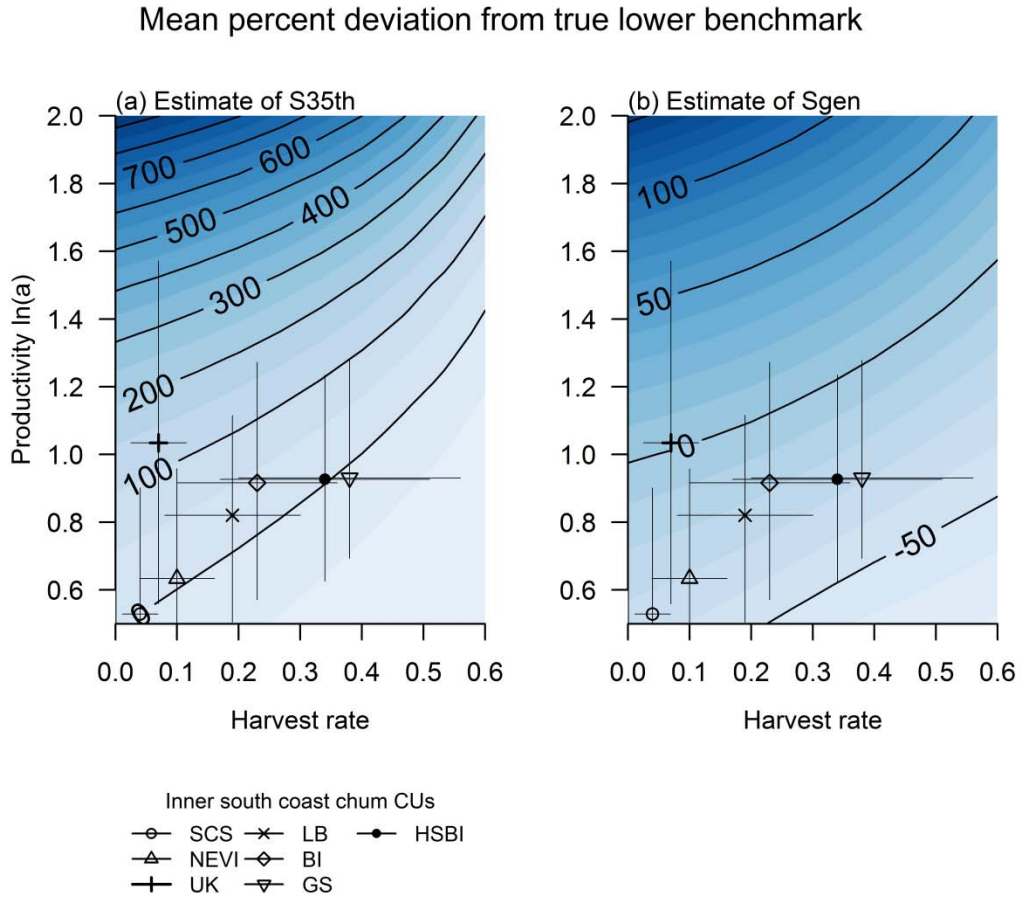


Fig. E3. Same as Fig. E1 except panel (a) depicts performance of the lower benchmark based on the 40<sup>th</sup> percentile of the observed spawner time-series. Panel (b) is the same as in Fig. E1, but is shown here for comparison.

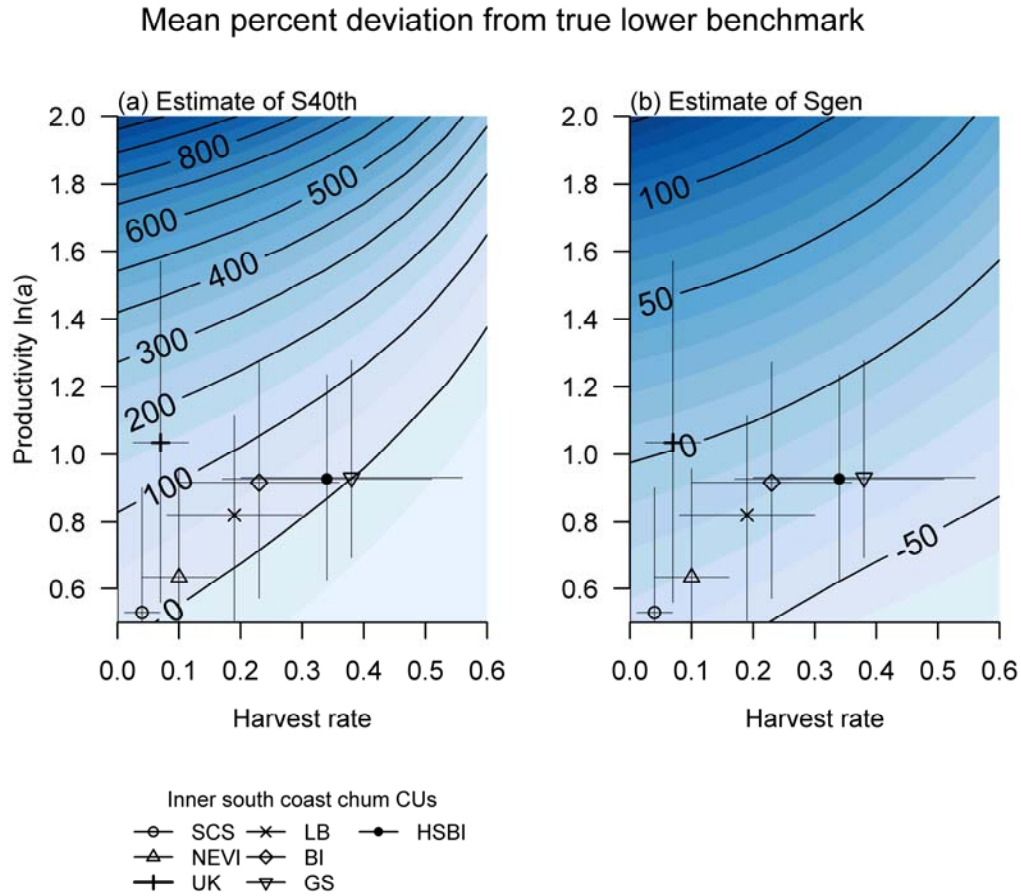


Fig. E4. Same as Fig. E1 except panel (a) depicts performance of the lower benchmark based on the 45<sup>th</sup> percentile of the observed spawner time-series. Panel (b) is the same as in Fig. E1, but is shown here for comparison.

