Fisheries and Oceans Canada
Pêches et Océans Canada

Ecosystems and Oceans Science
Sciences des écosystèmes et des océans

Evaluating Benchmarks of Biological Status for Data-limited Conservation Units of Pacific Salmon, Focusing on Chum Salmon in Southern BC

Carrie Holt[1], Brooke Davis[2], Diana Dobson[3], Lyse Godbout[1], Wilf Luedke[3], Joe Tadey[2], and Pieter Van Will[4]

[1]Fisheries and Oceans Canada
Pacific Biological Station
3190 Hammond Bay Road
Nanaimo, BC V9T 6N7

[2]Fisheries and Oceans Canada
100 Annacis Parkway
New Westminster, BC V3M 6A2

[3]Fisheries and Oceans Canada
3225 Stephenson Point Road
Nanaimo, BC V9T 1K3

[4]Fisheries and Oceans Canada
PO Box 10, 8585 Wollason Road
Port Hardy, BC V0N 2P0

September 2018

Canada

Foreword

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Correct citation for this publication:

Holt, C.A., Davis, B., Dobson, D., Godbout, L., Luedke, W., Tadey, J., and Van Will, P. 2018. Evaluating Benchmarks of Biological Status for Data-limited Conservation Units of Pacific Salmon, Focusing on Chum Salmon in Southern BC. DFO Can. Sci. Advis. Sec. Res. Doc. 2018/011. ix + 77 p.

Aussi disponible en français :

Holt, C.A., Davis, B., Dobson, D., Godbout, L., Luedke, W., Tadey, J., et Van Will, P. 2018. Évaluation des points de référence biologiques pour les unités de conservation du saumon du Pacifique pour lesquelles les données sont limitées, mettant l'accent sur le saumon kéta du sud de la Colombie-Britannique. Secr. can. de consult. sci. du MPO. Doc. de rech. 2018/011. ix + 87 p.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Status assessments for Chum Salmon (*Oncorhynchus keta*) under the Wild Salmon Policy (WSP) have been limited, in part because recruitment time-series required to calculate stock-recruitment based benchmarks are not consistently available. Alternative benchmarks have been proposed for data-limited Conservation Units (CUs) using percentiles of the observed spawner abundance time-series. However, these benchmarks have not been evaluated against stock-recruitment benchmarks currently used to assess status on abundances for data-rich CUs. Our goals were to evaluate percentile-based benchmarks against stock-recruitment based benchmarks accounting for high uncertainties and possible biases in spawner abundances, catches, recruitment estimates, and age-at-maturity. We used two approaches to evaluate benchmarks based on a retrospective comparison through the historical record and a prospective simulation model under numerous hypothetical future scenarios. We demonstrate an approach for providing assessments that accounts for uncertainties in benchmarks, and provide advice on the applicability of percentile-based benchmarks for data-limited CUs of Chum Salmon relative to stock-recruitment benchmarks used for data-rich CUs. In general, our results support the application of percentile-based benchmarks for data-limited CUs of Chum Salmon when productivity is moderate to high (>2.5 recruits/spawner) and harvest rates are low to moderate (≤40%). However, we suggest further evaluation of percentile benchmarks (and the consideration of alternatives) when productivity is expected to be low and/or harvest rates high. Under these conditions, concurrent declines in abundances and percentile benchmarks can results in status assessments that are more optimistic than those from stock-recruitment benchmarks due to a shifting baseline.

1   INTRODUCTION

1.1   BACKGROUND AND OBJECTIVES

Canada's Wild Salmon Policy (WSP) outlines strategies and action steps to restore and maintain healthy and diverse salmon populations for future generations (DFO 2005). In the first strategy of the WSP, benchmarks of biological status are used to assign status of Conservation Units, CUs (population units of biological assessment under the WSP), into one of three zones: green, amber and red, representing increasing conservation concern and management intervention. Benchmarks on abundances for data-rich CUs, defined here as CUs with time-series of spawners and recruitment, have been developed from stock-recruitment relationships (Holt et al. 2009) and applied to status assessments under the WSP for Sockeye Salmon (Grant and Pestal 2013), Coho Salmon (DFO 2015), and Chinook Salmon (DFO 2016).

The lower benchmark, delineating red and amber zones, is established at a level to ensure a buffer between it and a level 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. For data-rich CUs, the lower benchmark on abundances is set at, S_{gen}, the number of spawners required to recover to S_{MSY} (spawners at MSY) within one generation, under equilibrium conditions in the absence of fishing (Holt et al. 2009). The upper benchmark, delineating amber and green zones is set at the escapement level associated with the maximum average annual catch, under current environmental conditions. For data-rich CUs, the upper benchmark is 80% of S_{MSY} (Holt et al. 2009). While the 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 in BC. 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 Chum Salmon in southern BC, many CUs do not have reliable time-series of recruitment, though relative escapement time-series are often available. These CUs are considered data-limited for the purposes of this paper. Godbout et al. (2004) assessed status of data-limited Chum Salmon Fishery Management Areas in central and southern BC by comparing current abundances to historical medians. For data-limited salmon populations in Alaska, percentile-based approaches have been used to determine sustainable escapement goals (SEGs) by the Alaska Department of Fish and Game (ADFG), and have been proposed as biological benchmarks under Canada's Wild Salmon Policy (Holt and Folkes 2015). This type of benchmark requires escapement data only, and simply compares current escapement levels with the percentiles of historical observations, similar to metrics and benchmarks on long-term trends in spawner abundances already considered under the WSP (Holt et al. 2009). Data deficient CUs have been defined in previous WSP assessments as those without sufficient data to assign an overall status, missing both reliable time-series of spawner abundances and recruitment (Grant and Pestal 2013; DFO 2016).

Status assessments under the WSP integrate numerous metrics, including those on abundances, trends in abundance, and spatial distribution (Holt et al. 2009). Benchmarks on abundances (percentile or stock-recruitment based) are only one component of an integrated assessment of status that includes numerous other metrics (Grant and Pestal 2013; DFO 2015; DFO 2016).

Biological benchmarks under the WSP differ from management reference points that inform harvest decision rules, as biological benchmarks depend only on biological criteria and do not include socio-economic considerations (Holt and Irvine 2013). Management reference points are quantitative states that characterize desirable biological and/or economic properties of

fisheries, and can be used as a basis for harvest decision rules. Management reference points are required for the management of Chum Salmon domestically and internationally. For example, the Pacific Salmon Treaty (PST) Chum Annex requires management reference points for PST related fisheries, including the Johnstone Strait fisheries and subsequent terminal fisheries. Management reference points for Chum Salmon in southern BC are 20-35 years out of date, and do not reflect current trends in productivity, stock status, or other ecosystem considerations.

Marine Stewardship Council certification of southern BC Chum Salmon also requires assessment of status against reference points. Although Portley and Geiger (2014) suggested that WSP benchmarks (S_{gen} and 80% of S_{MSY}) were compliant with MSC certification requirements for reference points, data limitations have prevented their application in many cases. Instead, percentile-based benchmarks have been applied for MSC purposes (e.g., Hilborn et al. 2013, English et al. 2014) without a thorough evaluation of the properties and possible biases of those benchmarks compared with WSP benchmarks.

In one exception, the ADFG evaluated the extent to which various percentiles of spawner abundances approximated S_{MSY} as a basis for Sustainable Escapement Goals, SEGs (Clark et al. 2014). Based on this study, a multi-tiered system was recommended for escapement goals in Alaska, where percentile values are adjusted based on data contrast, data uncertainty, and harvest rates. However, the extent to which percentile benchmarks are consistent with biological benchmarks already identified under the WSP remains uncertain. For data-limited Chum Salmon CUs in southern BC, percentile benchmarks at the 25th and 75th percentiles have been proposed and provisionally implemented as lower and upper benchmarks, respectively, for MSC certification (Hilborn et al. 2013).

It is widely recommended that assessment methods be evaluated prior to implementation to avoid costly investments in approaches that provide little information or do not achieve stated objectives in practice (Peterman 2004, Butterworth 2007). Our objectives were to fill this gap for percentile-based benchmarks by:

- Evaluating biological benchmarks for data-limited Conservation Units of Chum Salmon based on percentiles of observed abundances by comparing them to standard model-based benchmarks, accounting for high uncertainties and possible biases in spawner abundances, catches, recruitment estimates, and age-at-maturity.

- Examining and identifying uncertainties in the data, methods, and benchmarks. Developing and demonstrating an approach for providing assessments that accounts for those uncertainties.

- Providing advice on the applicability of percentile-based benchmarks for data-limited populations of Chum Salmon in southern BC.

To emphasize, our goal was to provide a quantitative evaluation of percentile-based benchmarks for application to Chum Salmon, and not to provide specific assessment advice on individual CUs.

Here we employed two approaches to evaluate performance of percentile-based benchmarks based on their performance in the historical record and in a prospective simulation under numerous hypothetical future scenarios. In retrospective analyses, performance of benchmarks was evaluated by comparing temporal pattern of assessments using benchmarks derived from percentiles against those from stock-recruitment models. Status in red, amber or green zones was determined by comparing benchmarks to the generational (geometric) mean spawner abundances. In this way, we evaluated if biological status derived from data-limited benchmarks would have differed from those based on stock-recruitment benchmarks for data-rich CUs using

only data prior to the generation being assessed (Figure 1). These analyses were performed on a subset of CUs where both data-limited and data-rich benchmarks could be estimated.

Retrospective Analyses
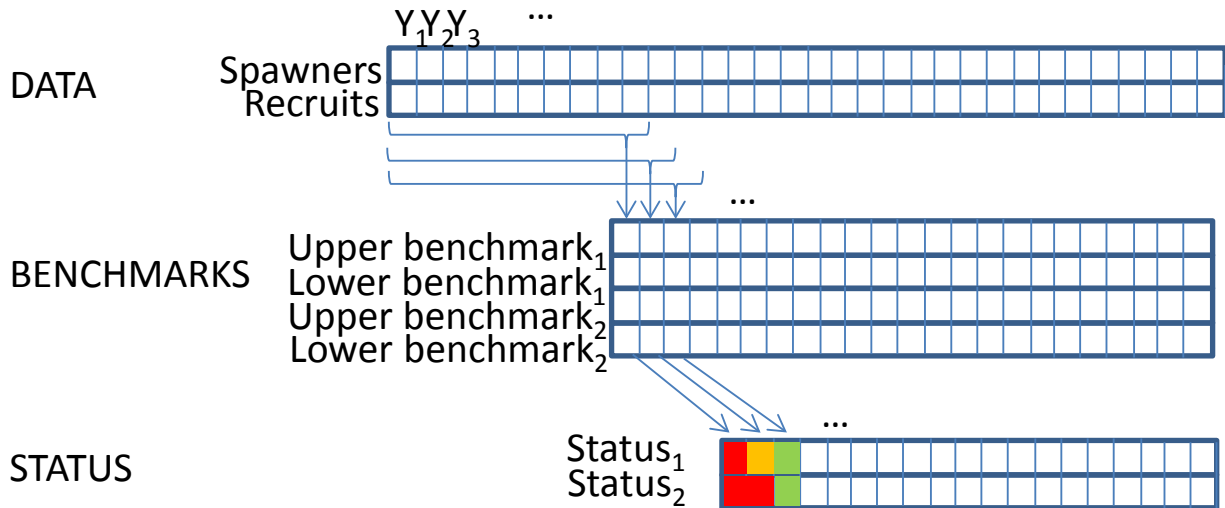
Figure 1. Schematic of retrospective analyses. A minimum of 10 years of spawner data or 10 years of paired spawner and recruitment data aligned by brood year are required to estimate percentile and stock-recruitment benchmarks, respectively. Y_1, Y_2, Y_3, represent brood years. Benchmarks derived from one of two methods are used to estimate statuses against generational mean abundances (geometric mean) 4 years later.

Simulation modelling has the advantage over retrospective analyses of evaluating benchmarks over a wide range of hypothetical future scenarios and explicitly accounting for uncertainty in data and population dynamics. Here, candidate benchmarks for data-limited CUs were evaluated against data-rich counterparts using a Monte Carlo simulation model that included stochastic variability in population dynamics, observations of abundances, and harvest. The simulation model employed a retrospective analysis as in Figure 1, but the underlying data were simulated instead of empirical, and analyses were repeated over numerous random iterations and scenarios.

To compare benchmarks using current methodologies, we evaluated percentile benchmarks against two alternate stock-recruitment benchmarks, one based on a standard, Ricker (1975), stock-recruitment model and one based on hierarchical Ricker stock-recruitment model that used information on productivity from neighbouring CUs to improve parameter estimates and reduce uncertainties (described in more detail in Section 2). Although a comparison between standard and hierarchical models is outside the scope of the current study, hierarchical models were included in our analyses because they are increasingly being applied to status assessment of Pacific salmon in BC where recruitment time-series are available (Dorner et al. 2013, Korman and English 2013).

Chum salmon generally spawn in the lower reaches of coastal streams or rivers, and their fry spend little time in freshwater (≤ 6 months). Outmigration occurs in spring but fry remain in estuaries throughout the summer before migrating offshore in the fall. Adults return to spawn generally at ages 3-5, in late summer and fall. Due to their limited jumping ability, the distribution of spawning is restricted to lower reaches of rivers with a few exceptions (e.g. Yukon River). Females select spawning habitat according to water depth, velocity, gravel composition, cover

and related characteristics (Groot and Margolis 1991). Density dependent survival may occur at spawning, freshwater and estuary rearing stages, and marine life-stages, where Chum Salmon compete for prey resources with other CUs and salmon species (Groot and Margolis 1991, Debertin et al. 2016).

2 EVALUATION OF PERCENTILE-BASED BENCHMARKS AGAINST DATA-RICH STOCK-RECRUITMENT-BASED BENCHMARKS USING RETROSPECTIVE ANALYSIS

The goal of our retrospective analysis was to compare status reached under both data-rich and data-limited biological benchmarks for Conservation Units in southern BC (Figure 2). For the data-rich scenarios, we compared benchmarks derived from two different forms of the Ricker stock-recruitment model: the standard model, which estimates parameters independently for each CU, and a hierarchical model where CUs within groupings are assumed to have productivity values from a shared distribution, centered on an overall mean productivity (Figure 3a). We used the Ricker model because it is the basis for abundance-based benchmarks under the WSP (Holt 2009, Holt et al. 2009) and it adequately captures stock-recruitment dynamics for Chum Salmon (Dorner et al. 2008). Given evidence for spatial covariation in productivities among Chum Salmon populations within regions in BC (Pyper et al. 2002), hierarchical models may reduce uncertainties and biases in parameter estimation by sharing information on productivity across populations. We used two regional groups for the hierarchical model: the Inner South Coast (including east coast of Vancouver Island, Strait of Georgia and Johnstone Strait) and the west coast of Vancouver Island, which aligned with different large-scale marine circulation systems (Holtby and Ciruna 2007). Given evidence for widespread changes in productivity over time (Malick and Cox 2016), we further identified temporal trends in productivity. This variability in productivity informed the magnitude of uncertainty in productivity to consider in scenarios generated within the simulation evaluation (Section 3).
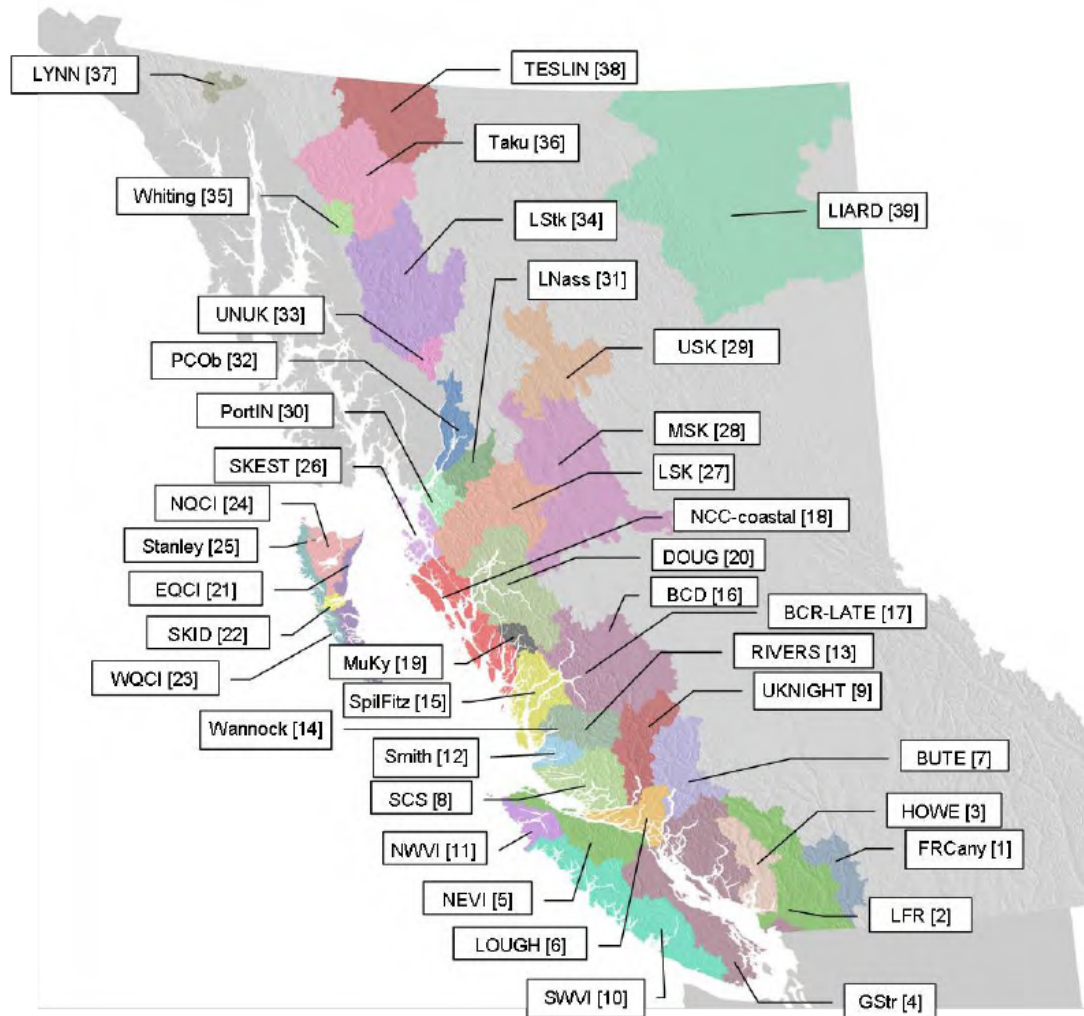
Figure 2. Map of Conservation Units for Chum Salmon in BC from Holtby and Ciruna (2007). The inner south coast CUs, labelled UKnight (Upper Knight), Bute (Bute Inlet), Howe (Howe Sound-Burrard Inlet) , GStr (Georgia Strait), Lough (Loughborough), SCS (Southern Coastal Streams), and NEVI (Northeast Vancouver Island), and west coast of Vancouver Island CUs, labelled SWVI (Southwest Vancouver Island) and NWVI (Northwest Vancouver Island), were included in our retrospective analyses.

(a)

Standard Ricker Model

Hierarchical Ricker Model

Priors

Parameters

Data

$\alpha_i \sim N(1,1)$ $\beta_i \sim unif(0, \max(S_i))$

$\mu_\alpha \sim N(1,1)$ $\sigma_\alpha \sim unif(0,100)$ $\beta_i \sim unif(0, \max(S_i))$

μ_α σ_α

α_1 α_2 ... α_n β_1 β_2 ... β_n

α_1 α_2 ... α_n β_1 β_2 ... β_n

S_1	R_1
s	r
s	r
...	...
s	r

S_2	R_2
s	r
s	r
...	...
s	r

S_n	R_n
s	r
s	r
...	...
s	r

CU_1 CU_2 CU_n

Ricker Parameter Estimates Across Models

Southern Coastal Streams — North East Vancouver Island — Upper Knight — Loughborough — Bute Inlet — Georgia Strait — Howe Sound to Burrard Inlet — North West Vancouver Island — South West Vancouver Island

Ricker α Estimate

S_{max} Estimate (1000's)

Std. Hier.

Ricker Model Form

Figure 3. (a) Schematic diagram of standard and hierarchical Ricker models. For the standard Ricker model, CU-specific parameters α and β (1/S$_{max}$) are estimated from uninformative priors. For the hierarchical model, priors on μ$_\alpha$ and σ$_\alpha$ are used to estimate a hyper-distribution of α parameters, from which CU-specific estimates are drawn. (b) Parameter estimates from the standard and hierarchical Ricker models (black and grey, respectively) in the final year (2012 for ISC CUs, 2015 for WCVI CUs) across CUs. Circles indicate posterior medians, and lines indicate 95% credible intervals.

2.1   DATA

2.1.1   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: Upper Knight, Bute Inlet, Howe Sound-Burrard Inlet, Georgia Strait, Loughborough, Southern Coastal Streams, and Northeast Vancouver Island. Escapement data, identified as either wild-origin (naturally spawning) or hatchery-origin, were available for these CUs from 1953-2012, while CU-specific return data were reconstructed from catches, migration timing and patterns, spawner abundances, and age distributions, for brood years 1955-2006 (Van Will 2014). Wild recruitment to systems with associated hatchery stocks were estimated by assuming that proportions of wild fish in catches were equal to the proportion in observed escapement. Historical wild and hatchery composition of the catches was not available to identify the respective proportion of catches, though hatchery and wild stocks are thought to be harvested at approximately equal rates. Exploitation rate time-series for all CUs examined here are provided in Appendix A.

In years when spawner abundances (escapement) were missing, data have been infilled using a commonly applied approach that assumes 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. Although infilling across CUs reduces independence of data among CUs, it was done infrequently and is unlikely to cause a consistent bias in our results. Infilled escapement and return data were combined with age-composition data to create brood tables from which stock-recruitment time series were 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), thereby reducing these possible biases (Walters and Martell 2004). However, these results should be considered in conjunction with those from simulation model that incorporates multiple sources of data and assessment uncertainties.

2.1.2   West Coast of Vancouver Island (WCVI) Chum Salmon Data

Time series of escapement and returns from 1953-2015 were available for both the Southwest and Northwest coast of Vancouver Island CUs (SWVI, NWVI). Brood tables were constructed based on yearly age composition data beginning in 1959, but with a gap in the mid-1960's. The corresponding stock-recruitment time-series ranged 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. Specifically, hatchery contributions in Nitinat (Area 22) and Tlupana (Area 25) within the SWVI CU 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 and catches from this Area were removed. Similar to the ISC data set described above, the WCVI CUs 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 the run reconstruction stock-recruit time

series should not be considered the "true" state of the population, and these results should be considered alongside simulation results, which consider these uncertainties.

2.2 METHODS

For both stock-recruitment based and percentile-based 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). 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 CUs, and 1957-1958, 1965/1966-2010, for WCVI CUs. 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.

2.2.1 Standard Ricker Model

For each year with sufficient data and for each CU i, a standard Ricker model (Eqn. 1) was fit in a Bayesian context, using Markov Chain Monte Carlo (MCMC) methods, using all data available up that year (brood years $y = 1\ to\ Y$).

(1) $R_{i,y} = \alpha_i S_{i,y} e^{-\beta_i S_{i,y}}$,

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 α (also referred to as productivity) is recruits-per-spawner at low spawner abundances, and β 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 τ_v represents precision of process error and precision is the reciprocal of variance.

(2) $\log(R_{i,y}) = \log(\alpha_i) + \log(S_{i,y}) - \beta_i S_{i,y} + v_i,\ v_i \sim normal(0,\ 1/\tau_{v_i})$.

We put a weakly informative prior on α to ensure values greater than zero and within the bounds of observed productivity values for Chum Salmon (Dorner et al. 2008) (See Appendix B for plots of priors and posteriors of α parameter).

(3) $\log(\alpha_i) \sim normal(1,1)$.

The prior for β was set indirectly by applying a prior on its reciprocal, S_{Max}. We had no independent prior information on S_{Max} for these CUs, so we applied a uniform distribution bounded by 1 and twice the maximum observed spawner value (Eqn. 4). In a sensitivity analyses, we also considered a diffuse log-normal distribution for the prior (see Appendix C).

(4) $S_{\max_i} \sim uniform(1, max(S_{obs_i}) \cdot 2)$

Uninformative gamma priors were used for τ parameters,

(5) $\tau_{v_i} \sim gamma(0.01, 0.001)$.

Although other forms of the Ricker model with environmental co-variates have been evaluated for Chum Salmon (Godbout et al. 2006; Malick et al. 2017), they have not yet been used to

develop WSP benchmarks. A full evaluation of data-rich benchmarks was beyond the scope of the analyses, and those models were not considered here.

2.2.2 Hierarchical Ricker Model

We also 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 α_i values were drawn from a common, normal distribution (Figure 3a, right side),

$$(6)\ R_{i,y} = \alpha_i S_{i,y} e^{-\beta_i S_{i,y}} e^{v_i}, v_i \sim normal\left(0, 1/\tau_{v_i}\right), \alpha_i \sim normal(\mu_\alpha,\ 1/\tau_\alpha),$$

where μ_α is the mean of the normal distribution and τ_α is precision (the reciprocal of variance).

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, μ_α.

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

We put an uninformative prior on the variance, σ_α^2, where $\sigma_\alpha^2 = 1/\tau_\alpha$,

$$(8)\ \sigma_\alpha^2 \sim Uniform(0, 100).$$

Models were fit using MCMC runs using JAGS (Plummer 2003) interfaced through R version 3.2.0 (R Development Core Team 2017) using package "R2jags" (Su and Yajima 2012). Model convergence was assessed using Gelman-Rubin and visual inspection of trace plots. Each model run included three independent MCMC chains of length 500,000, with a burn in of 200,000 and thinning rate of 300. Gelman-Rubin statistics compare within-chain to between-chain variance, to ensure that all chains are converging on the same solution. Chains are generally considered converged when Gelman-Rubin statistics are below 1.1 (though final year models were below 1.05).Trace plots have not been included here due to the large number of models considered in the retrospective analysis.

2.2.3 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),

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

Percentile benchmarks were calculated as the 25th and 75th percentile of observed spawner abundances (escapement) ranked from lowest to highest, for the lower and upper benchmarks respectively (S_{25th}, S_{75th}). We used these percentiles as they have been used for Chum Salmon in BC (Hilborn et al. 2013). Holt and Ogden (2013) recommended against using stock-recruitment benchmarks when Ricker α falls below 1.5 because at low productivity the lower benchmark, S_{gen} tends to be higher than 80% of S_{MSY}, the upper benchmark. We have removed years when α <1.5 from our retrospective analysis.

2.2.4 Retrospective Analyses

Retrospective analyses were implemented by annually estimating benchmarks and assessing status using all available data up until that year. This approach mimics the benchmarks and statuses that would have been generated historically. We derived bootstrapped confidence intervals for percentile-based benchmarks by resampling the time-series with replacement to generate a distribution of lower and upper benchmarks. We used simple non-parametric bootstrapping which may over-estimate confidence intervals if time-series are autocorrelated. Methods that account for temporal autocorrelation should be considered when applying this method in the future. The current analyses are for illustrative purposes only. Similarly, we derived Bayesian credible intervals for stock-recruitment based benchmarks by estimating lower and upper benchmarks for each MCMC sample of the posterior distribution of Ricker parameters. We expressed uncertainty in benchmarks as the 95% credible intervals, estimated as the 2.5[th] and 97.5[th] posterior densities of each benchmark.

We used two approaches to characterize the uncertainty in the resulting status assessments (i.e., assigning red, amber, or green). First we assessed status using the lower and upper credible intervals of benchmarks to capture the maximum range of statuses (as shown in transparent bars in Fig. 6). Second, to account for the correlation between upper and lower benchmarks (e.g., due to the correlation between Ricker parameters), we also assessed status using the individual sets of lower and upper benchmarks from each bootstrapped time-series for the percentile-based benchmarks, and for each MCMC trial of the Bayesian analysis for the stock-recruitment benchmarks. From the resulting distribution of statuses, we generated a probability associated with each status (red, amber, or green) for both types of benchmarks (as shown by vertical bars in Fig. 7).

2.2.5 Changes in Productivity

Finally, 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 each stock-recruitment time-series, which allowed α to vary over time for each CU individually (Grant et al. 2011). We fit this model using all available data for each site. The model is similar to the standard Ricker form, but with a time-varying α parameter,

$$(11) \qquad R_{i,y} = \alpha_{i,y} S_{i,y} e^{-\beta_i S_{i,y}} e^{v_i}, \ v_i \sim normal(0, 1/\tau_{v_i}),$$

where $\alpha_{i,y}$ is productivity in brood year y, for CU i. The model assumes that α changes over time following a Gaussian random walk,

$$(12) \qquad \log(\alpha_{i,y}) = \log(\alpha_{i,y-1}) + w, \ w \sim normal(0, 1/\tau_{w_i}).$$

The same prior distributions were applied as for the standard Ricker model (Eqns. 3-5), with the addition of a normally distributed prior on α in year 1, and a uniform prior on the variance associated with the Gaussian random walk $\sigma_{w_i}^2$, where $\sigma_{w_i}^2 = 1/\tau_{w_i}$,

$$(13) \qquad \log(\alpha_{i,1}) \sim normal(1,1), \text{ and } \ \sigma_{w_i}^2 \sim Uniform(0,100).$$

Convergence was assessed as described above, with no parameters having Gelman-Rubin statistics above 1.05.

2.3 RESULTS

2.3.1 Current Benchmarks and Status

Model fit for standard and hierarchical Ricker models varied by CU, ranging from moderate to very poor (r^2 between predicted and observed 0.41-0, Appendix D). We used all CUs regardless of model fit in our initial retrospectives analysis, but then considered only those CUs with relatively high correlations (r^2 values >0.25), Southern Coastal Streams and Northeast Vancouver Island in a subsequent analysis.

The relationship between the lower percentile benchmark, S_{25th}, and Ricker-based benchmark, S_{gen}, in the most recent year depended on the CU. The lower percentile benchmark was lower than S_{gen} for 4 CUs (Southern Coastal Streams, Upper Knight, Bute Inlet, and Howe Sound to Burrard Inlet), higher than S_{gen} for 4 CUs (Loughborough, Georgia Strait, Northwest Vancouver Island, and Southwest Vancouver Island) and similar to S_{gen} for 1 CU (Northeast Vancouver Island) (Table 1). In contrast, upper percentile benchmarks (S_{75th}) were generally much higher than the Ricker-based upper benchmarks ($80\%S_{MSY}$) (Figure 4 for an example CU and Appendix D for remaining CUs, 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 (see Appendix E).

Final Year Ricker Curves and Benchmarks
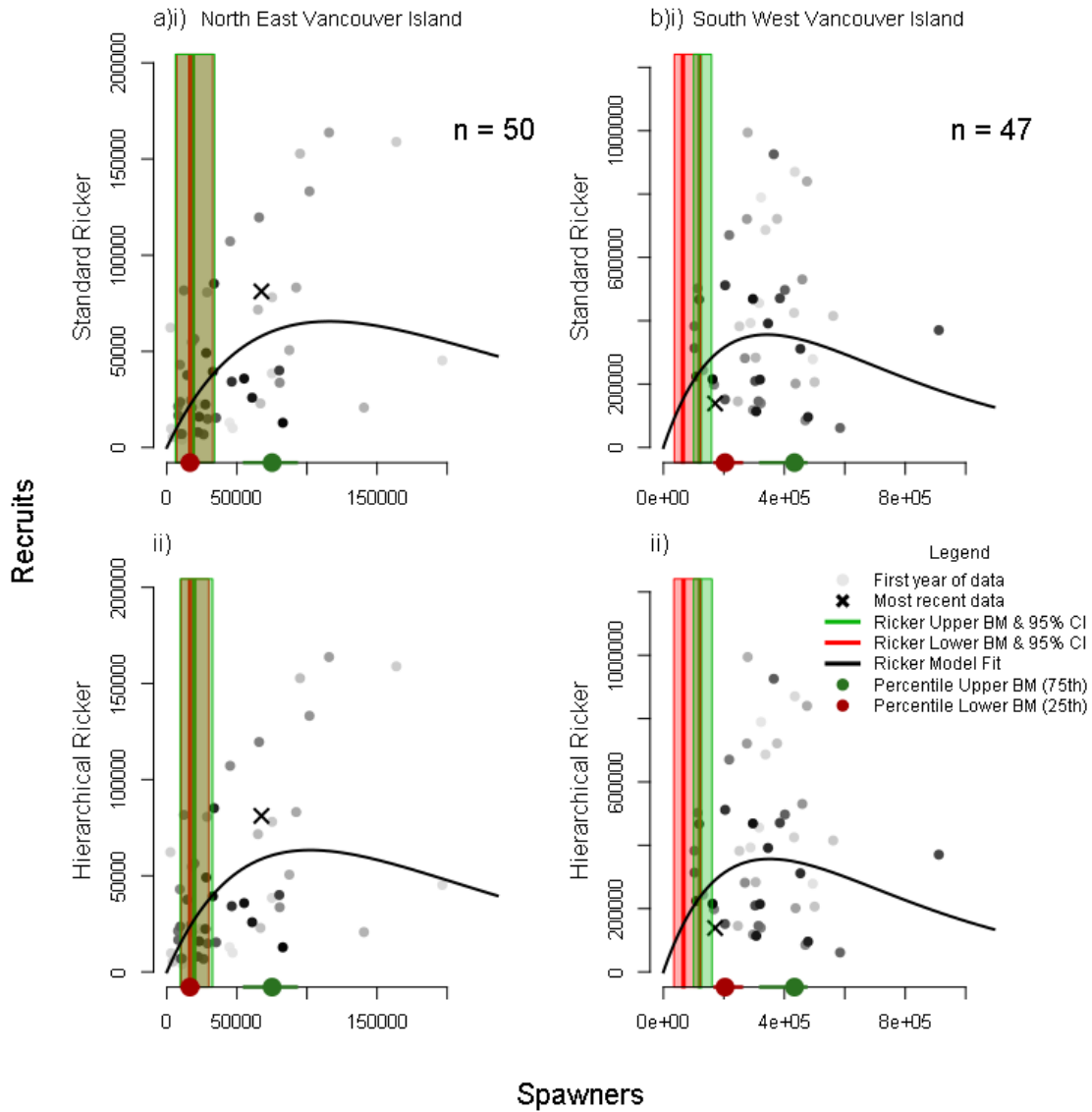
Figure 4. Observed spawner-recruit data with fitted Ricker curves and associated benchmarks for (i) the standard Bayesian Ricker model, and (ii) the Bayesian hierarchical Ricker model for selected CUs. Shaded regions indicate 95% credible intervals (CIs), delineated by 2.5th and 97.5th posterior densities. Red and green circles on x-axis identify percentile-based benchmarks with 95% CIs (S_{25th} and S_{75th}, respectively). Cross indicates most recent data point, for brood year 2006 for ISC CUs, 2010 for WCVI CUs. Points range from light grey (beginning of time-series) to black (end of the time-series).)Plots for all CUs can be found in Appendix D.

Stock-recruitment benchmarks varied slightly between the standard and hierarchical Ricker models (Tables 1 & 2, Figure 4, Appendix D), 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., Northeast Vancouver Island (Figure 4a).

Statuses for the most recent generation (ending in 2012 for ISC CUs, 2015 for WCVI CUs) were determined from benchmarks using all data available up to, but not including the most recent generation (Table 3). Both Ricker benchmarks yielded green status for all CUs, except for South Coast Streams; which was estimated to be in the red zone. Percentile-based statuses were the same or more precautionary for all CUs; only assessing green status for three out of nine CUs.

2.3.2 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 (Figure 5, bottom row compared to first two rows of panels). For WCVI CUs, which did not exhibit large contrast, percentile benchmarks were fairly consistent over time (Figure 5h and i).

Comparing the hierarchical model with the standard Ricker model, we found that uncertainties in estimates of α and S_{max} (Ricker parameters) were reduced slightly for the hierarchal model in some CUs on the Inner South Coast (Table 2, Figure 3b). The hierarchical Ricker model tended to reduce uncertainties for those CUs where productivity estimates were similar between the two models (comparing length of black and grey lines and location of black and grey points in Figure 3b, top row, e.g., Upper Knight). Alternatively, when productivity estimates from a CU differed from neighbouring CUs (and hence between standard and hierarchical Ricker models), uncertainty bounds stayed the same or increased (e.g., Northeast Vancouver Island, Figure 3b, top row).

Temporal Variability in benchmarks
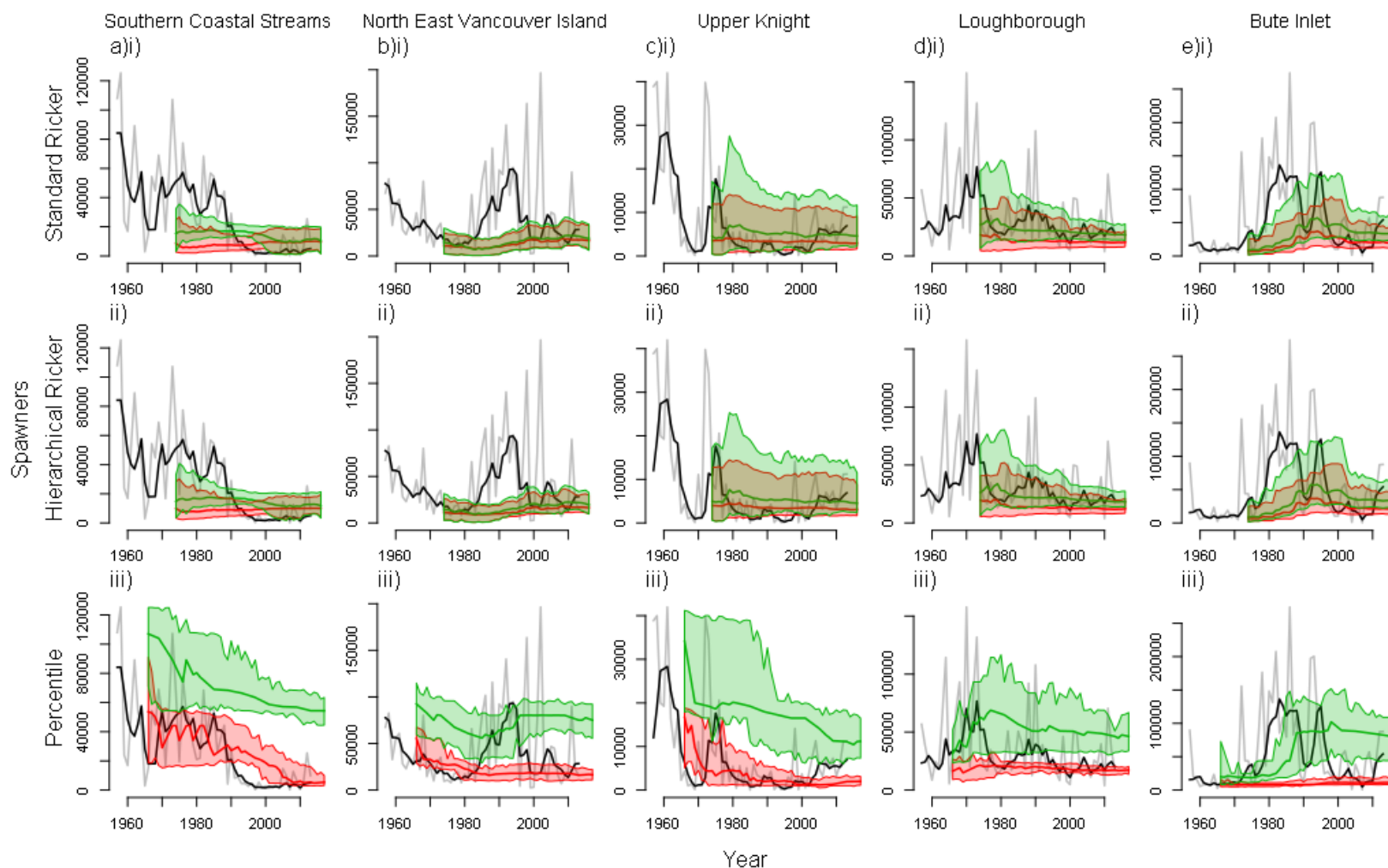
Figure 5. Raw (black lines) and generational average (grey lines) escapement over time for CUs, with retrospective estimates of upper (green line) and lower (red line) benchmarks, derived from (i) the standard Ricker model; (ii) a hierarchical Ricker model; and (iii) 25th and 75th percentiles. Shaded regions indicate 95% credible intervals. Retrospective benchmarks use all available data up to that generation.

Temporal Variability in benchmarks
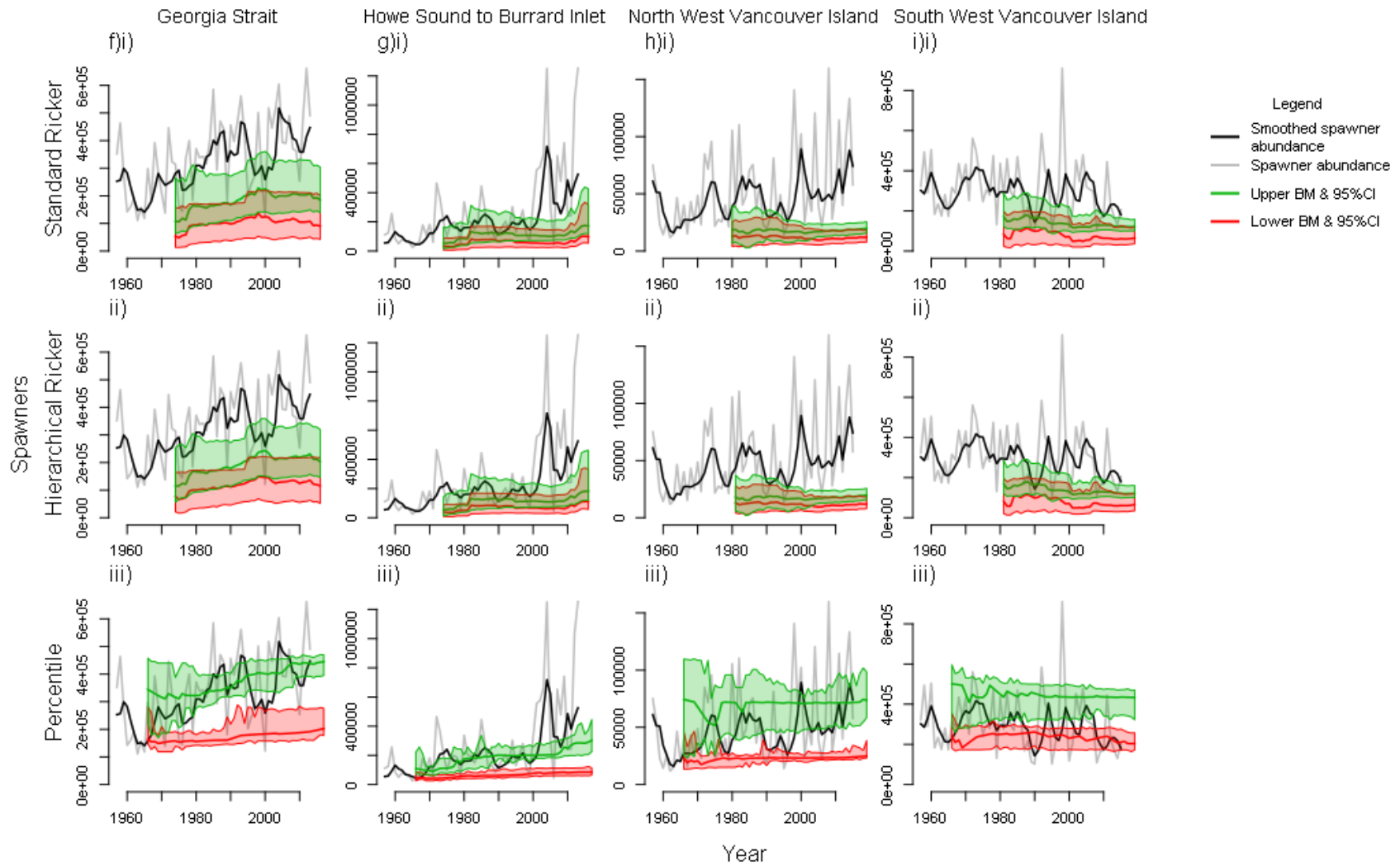
Figure 5 continued.

Stock-recruitment benchmarks tended to remain relatively consistent over time for four CUs (Upper Knight, Loughborough, Southwest Vancouver Island, Northwest Vancouver Island), exhibited divergent trends between upper and lower benchmarks (Southern Coastal Streams), or increased over time (Northeast 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 Figure 5). Uncertainties in stock-recruitment benchmarks declined over time for some CUs as longer time series improved precision, while others showed variable patterns over time possibly due to temporal variability in underlying stock-recruitment parameters.

Large uncertainties in stock-recruitment benchmarks resulted in uncertainties in status assessments, which we present in two ways. Status based on the lower and upper credible intervals of Ricker benchmarks are shown as light bars below and above the assessments from the median benchmark estimates in Figure 6. 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 (Figure 6b). The probability of a CU having each status, in each year, is depicted in Figure 7a).

2.3.3 Comparing Status across Benchmarks

The proportion of years where the percentile and Ricker-based benchmarks gave the same status varied across CUs, but averaged 43 and 44% for the standard Ricker and hierarchical Ricker models, respectively for ISC CUs (Table 4). Alternatively, for WCVI CUs, percentile and Ricker-based benchmarks rarely matched (never for Southwest Vancouver Island, 14-15% of years for Northwest Vancouver Island). On average, the percentile benchmark provided the same or lower status than Ricker-based benchmarks in 95% of years for both model types across CUs (i.e., being more precautionary because the application of relatively large percentile benchmarks sometimes indicated poor status when relatively small Ricker benchmarks indicated moderate to healthy status) (Table 4), despite percentile-based benchmarks occasionally dropping below Ricker-based benchmarks (Table 1 and Figure 5). The relatively few years when status from percentile benchmarks was higher than that from Ricker-based benchmarks (i.e., indicating healthier status and being less precautionary) were associated with long periods of both high and low abundances with abrupt transitions between them, resulting in relatively high Ricker benchmarks, but low percentile benchmarks (e.g., Upper Knight from 1999-2001, Figure 6c and 7c, and Bute Inlet in the early 2000s, Figure 6e and 7e).

The two Ricker-based benchmarks (standard and hierarchical Ricker) gave the same status 99% of years when averaging across CUs.

When we considered only CUs where models fits were relatively good (r^2>0.25, Southern Coastal Streams and Northeast Vancouver Island), we found that percentile-based benchmarks always matched or provided more precautionary status than Ricker-based benchmarks (Table 4)

Status Over Time

a) Southern Coastal Streams
b) North East Vancouver Island
c) Upper Knight
d) Loughborough
e) Bute Inlet
f) Georgia Strait
g) Howe Sound to Burrard Inlet
h) North West Vancouver Island
i) South West Vancouver Island

Legend
— Smoothed relative spawner abund.
— Relative spawner abundance
— Status
— Status 95% CI
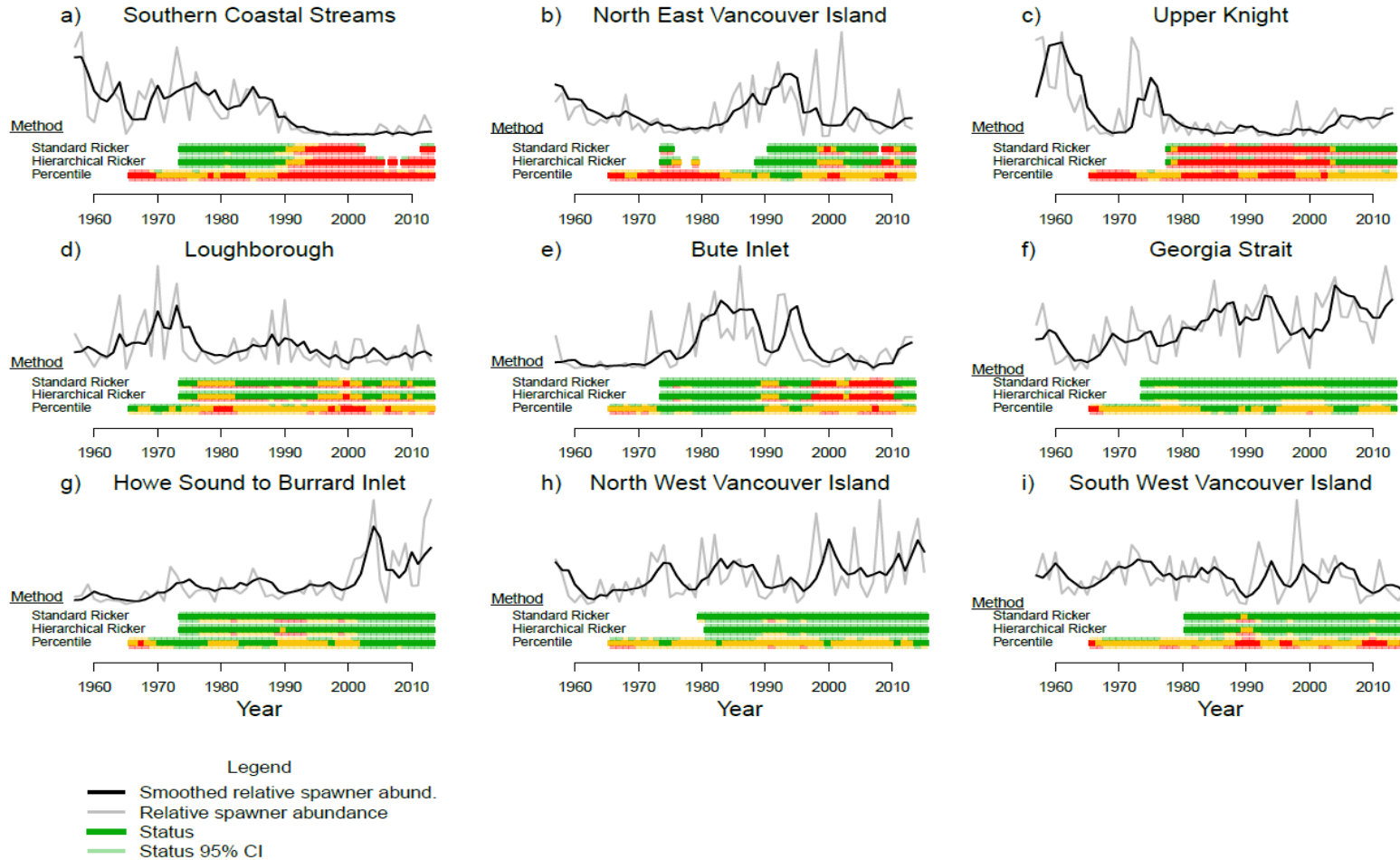
Figure 6. Standardized raw (grey) and generational geometric average (black) escapements across CUs, with status calculated from three methods indicated by coloured bars below. Transparent bars above and below the solid coloured bars for each benchmark indicate status at upper and lower 95% credible/confidence interval (CI) bounds for each benchmark and year. For Bayesian Ricker models, credible intervals were calculated as the 2.5th and 97.5th posterior densities for each Ricker-based benchmark. For percentile benchmarks, confidence intervals were calculated from a standard non-parametric bootstrap of the time-series. Gaps exist where data are missing, or when Ricker α drops below 1.5, at which point Ricker benchmarks are not recommended.

Probability of Status Over Time

a) Southern Coastal Streams
b) North East Vancouver Island
c) Upper Knight

d) Loughborough
e) Bute Inlet
f) Georgia Strait

g) Howe Sound to Burrard Inlet
h) North West Vancouver Island
i) South West Vancouver Island
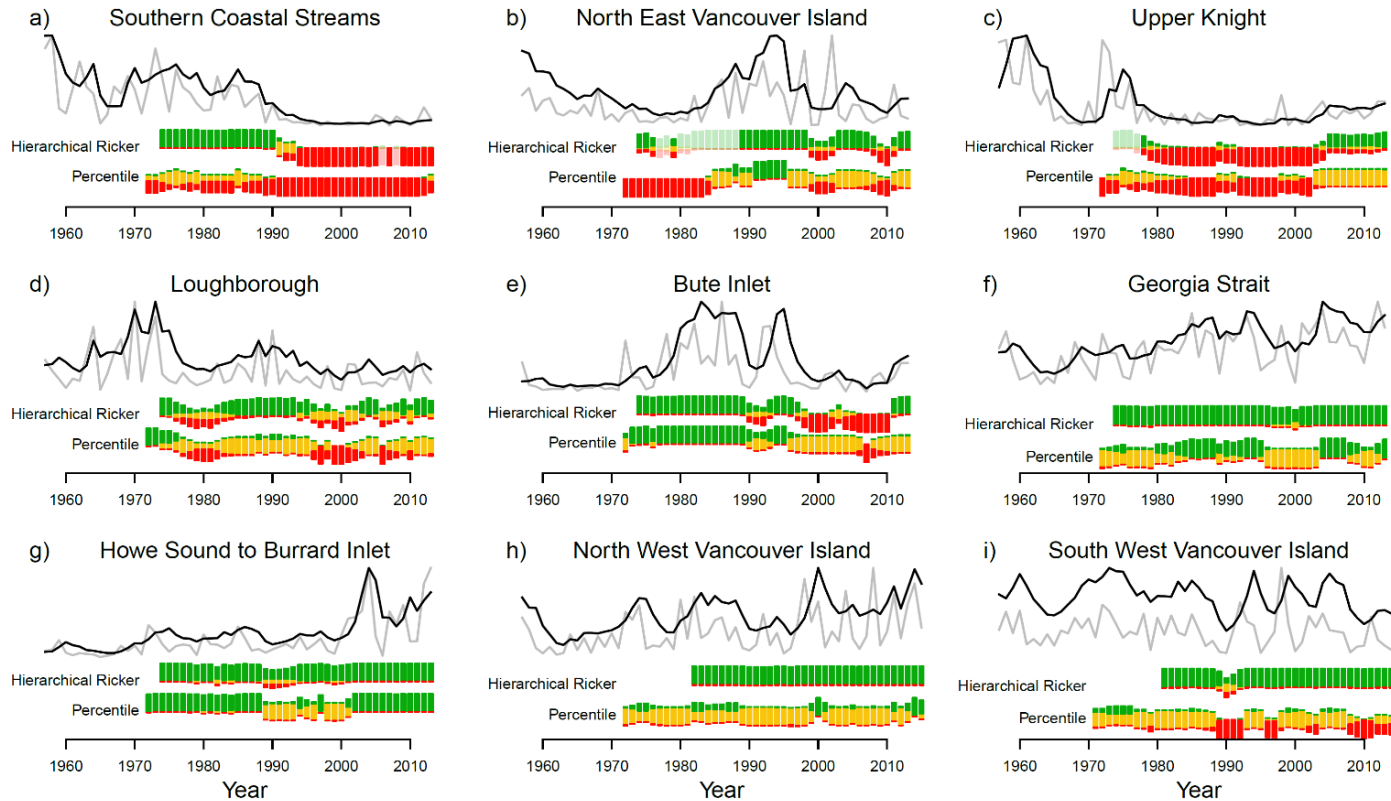
Legend
— Smoothed relative spawner abundance
— Relative spawner abundance
— Status

Figure 7. Standardized raw (grey) and generational geometric average (black) escapements across CUs, with status probabilities as assessed using hierarchical Ricker and percentile benchmarks, indicated by vertical coloured bars. The coloured proportions of each vertical bar represent the probability that status falls within each colour zone, red, amber and green. The vertical placement of the bars is centered around the amber zone; an upward shift represents a higher probability of green status, and a downward shift representing a higher probability of red status. Grey status bars (as in NEVI in the 1980's) represent years where Ricker α fell below 1.5, and Ricker benchmarks are not recommended.

2.3.4 Productivity over Time

Temporal patterns in productivity varied across CUs (Figure 8). Declines in productivity over time were observed in three CUs (Southern Coastal Streams, Loughborough, and Southwest Vancouver Island, Figure 8a,d, and h), increases followed by declines in three CUs (Northeast Vancouver Island, Bute Inlet, and Georgia Strait, Figure 8b, e, and f) and consistent levels followed by a small increase in Howe Sound to Burrard Inlet and Northwest Vancouver Island (Figure 8g and i). Estimates of productivity for Upper Knight (Figure 8c), were highly variable and uncertain. There was considerable uncertainty in productivity for all CUs, indicated by wide error bounds. Using data to estimate stock-recruitment parameters and benchmarks that spans decades where α has changed considerably may increase uncertainty in parameter estimates in the standard and hierarchical Ricker models.
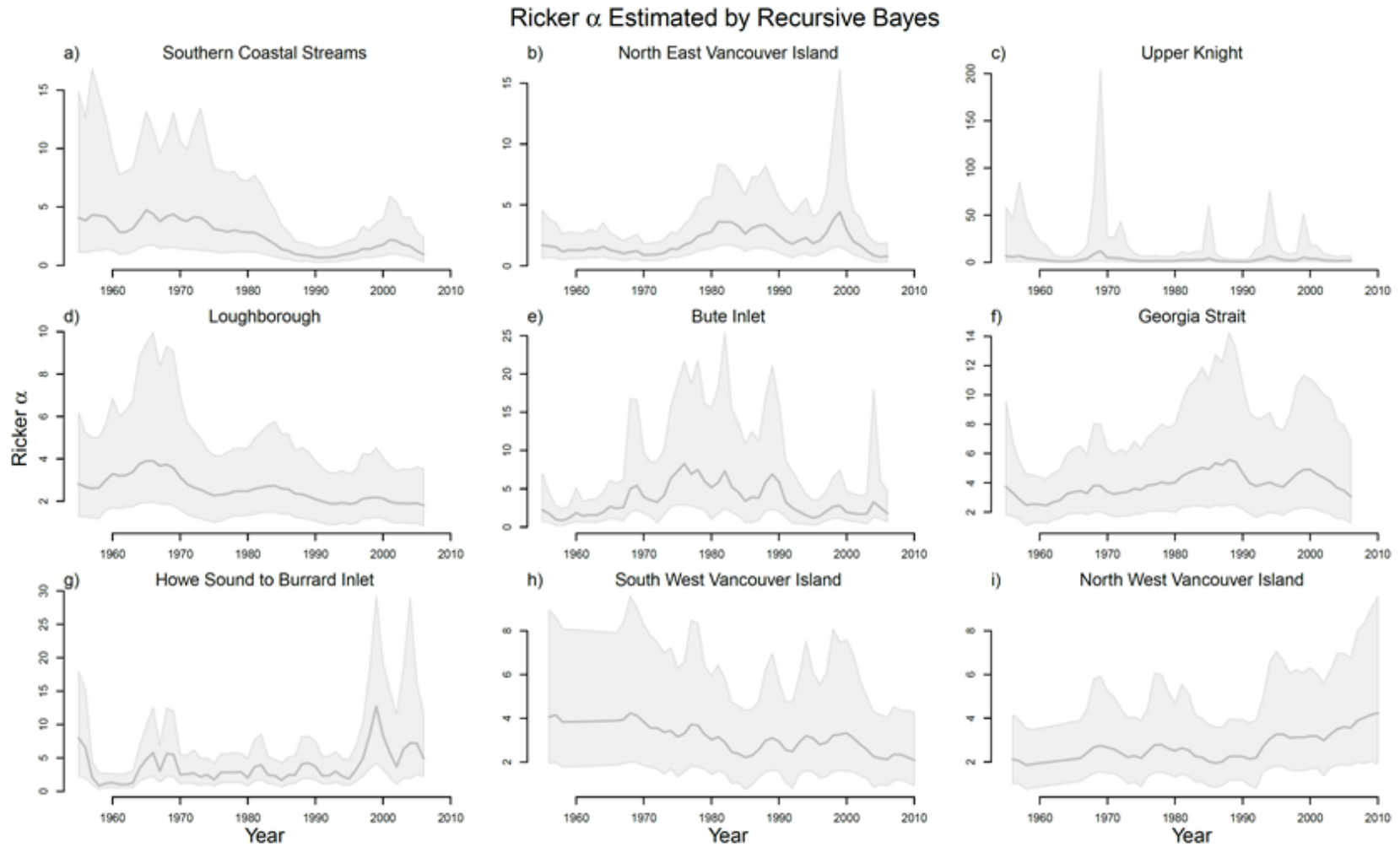
Figure 8. Estimated Ricker αvalues for CUs using a recursive Bayes model, which allows α to vary over time within a given CU. Grey shaded polygons indicate 95% credible intervals based on posterior distribution of estimated αvalues.

2.4 DISCUSSION

Our retrospective analyses indicate that the relationship between percentile-based lower benchmarks and Ricker-based lower benchmarks varies by CU and over time. Assessments from percentile-based lower benchmarks tended to be the same or lower than those from Ricker-based benchmarks. Upper 75th percentile benchmarks tended to be much higher than Ricker-based upper benchmarks. Percentile benchmarks yielded green status in only 18% of years across CUs, and 75% of years for Ricker benchmarks. Upper percentile benchmarks likely lie far above S_{MSY} levels. Overall, statuses based on percentile benchmarks were generally the same or lower than Ricker-based benchmarks adopted under the WSP. 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 relatively long-periods of low abundances with abrupt transitions to and from periods of high abundances.

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 99% of CU-year combinations. The CU-specific time-series were equally informative for the most part, so parameters changed little when combined in the hierarchical analyses. However, benchmarks derived from the hierarchical Ricker model were more certain than those from the standard model in cases where productivity was similar across CUs. Given large uncertainties in stock-recruitment data and inconsistent time-series for Chum Salmon in BC, a hierarchical approach is recommended over the standard Ricker model when there is support for the assumption of similar productivities among CUs.

3 EVALUATION OF PERCENTILE-BASED BENCHMARKS AGAINST RICKER BENCHMARKS USING SIMULATION MODELLING

3.1 METHODS

3.1.1 Model Description

The simulation model to evaluate benchmarks was structured in a similar way as the retrospective analysis described in Section 2 and Figure 1, except the underlying data were simulated instead of empirical, and analyses were repeated over numerous random iterations and scenarios. We adapted the simulation model of Holt and Folkes (2015) to compare estimates of data-limited and data-rich, lower and upper benchmarks against "true" underlying benchmarks, defined here as the "true" S_{gen} and 80% of S_{MSY}, respectively, derived using underlying known stock-recruitment parameters. As in Holt and Folkes (2015), the model included five components representing population dynamics, observations of abundances, assessment (including the derivation of benchmarks), harvest, and performance evaluation (Figure 9, Appendix F for model equations). The model emulated a single, generic CU of Chum Salmon with multiple sub-populations, where CU-level benchmarks were estimated by aggregating data across individual sub-populations.
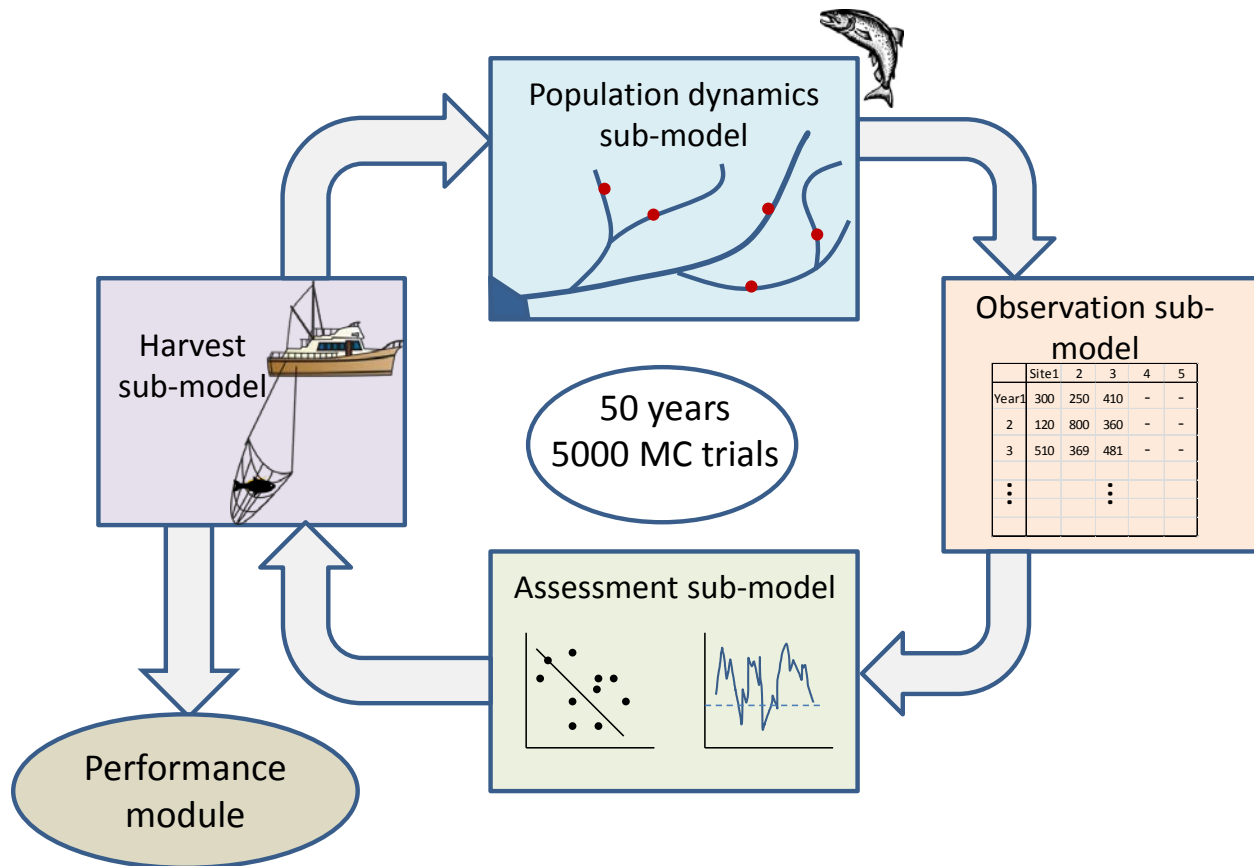
Figure 9. Schematic of simulation model used to evaluate benchmark performance.

In particular, the population sub-model included natural variability in adult recruitment based on a Ricker spawner-recruitment relationship, covariation in recruitment residuals among sub-populations, straying from the natal sub-population to neighbouring sub-populations when adults return to spawn, and variable age at maturity. Because data used to estimate benchmarks are often of poor quality, we included errors in observations of spawners, catch, and age-at-maturity in the observation sub-model. Benchmarks were estimated in the assessment sub-model from time-series of observed spawner abundances, estimated adult returns (observed catch plus observed spawner abundances), and derived recruitment aligned by brood year given observed ages at maturity. We also evaluated scenarios where only a portion of sub-populations 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. We further considered scenarios where no sub-populations were sampled in a given year, resulting in gaps in the observed time-series at various frequencies and a reduced length of the time-series.

The assessment sub-model model included the derivation of percentile-based benchmarks and stock-recruitment benchmarks applied under the WSP. Stock-recruitment benchmarks were estimated using a simple linear regression on the natural logarithm of the Ricker model (see Appendix F, Eqn. F16), a simplification of the Bayesian models executed in the retrospective analysis, required for computational efficiency. When prior distributions are uninformative, the parameters estimated with Bayesian methods approach those estimated using a linear regression. In preliminary analyses, we evaluated the performance of benchmarks derived from a hierarchical Bayesian model, an alternative data-rich benchmark, as described in more detail

in Section 2.2.2. We found that estimates of benchmarks from the hierarchical model were more accurate than from the standard model, but differences were very small (within 1%). Due to similarities in results and additional computational requirements of the hierarchical model, we focus on the standard Ricker model here.

The harvest sub-model included a target harvest rate with uncertainty in the outcomes from implementing the target due to, for example, non-compliance and variability in catchability (Holt et al. 2006).

In the performance module, benchmarks were evaluated based on the deviations between benchmark estimates and the "true" underlying values. For percentile benchmarks, we evaluated deviations between the 25th percentile (lower benchmark) and the "true" S_{gen} value, and between the 75th percentile (upper benchmark) and the "true" 80% of S_{MSY} value. Specifically, we evaluated, mean percent error, MPE, 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 this metric. In addition, this metric is scale independent facilitating comparisons across benchmarks and CUs. Preliminary results on mean raw error suggested patterns of results were similar across performance metrics.

The model was run over 50 years (plus 5 years pre-initialization and 20 years of initialization) and 5000 MC trials, the number of trials required to stabilize output metrics at (standard error ≤3% in performance metrics). The model was initialized for 20 years after a 5-year pre-initialization period necessary to generate the first year of recruitment. The model was developed in R version 3.2.2 (R Development Core Team 2017).

3.1.2 Model 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 Salmon were not available. The productivity parameter of the spawner-recruitment relationship, α, defined as recruits/spawner at low spawner abundance and referred to simply as productivity here, and the range considered in sensitivity analyses (Appendix F, Table 5) were chosen to bound productivities observed for six Chum salmon stocks from across BC (Dorner et al. 2008; ranging from 2.69-6.96), three stocks in the Skeena watershed, BC (Korman and English 2013; ranging from 2.01-2.86), and 9 CUs of Chum Salmon examined here (Figure 9). In preliminary analyses, we also considered temporal variability in productivity on benchmark performance, in the form of a step-like regime shift from moderate to low (high) productivity. However, we found that benchmark performance was similar to the scenario of constant low (high) productivity but with reduced magnitude of effects. Therefore, we have limited our analyses here to constant productivity over a range of plausible values. Productivities and spawner abundances at equilibrium abundances, S_{eq} (set at 10 000 fish) were assumed to be equal among sub-populations.

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 and English 2013), and considered a range of plausible autocorrelation coefficients (0 and 0.9) in sensitivity analyses. The standard deviation in recruitment residuals (in log-space) was set to 0.75, within the range of values estimated for the standard Ricker model (Section 2.3, 0.25-1.42), and previously for Chum Salmon on the Skeena River, BC (0.68-0.90) (Korman and English 2013), and within the range estimated for Sockeye Salmon in BC and Alaska (Korman et al. 1995, Peterman et al. 2003). The average proportions of mature adults at ages 3, 4, and 5 were estimated for 22 Chum Salmon stocks in BC and Alaska (0.24, 0.67, and 0.09, respectively, Pyper et al. 2002). The variance in the proportions of ages at maturity was estimated from empirical time series of

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 of adult recruits among all sub-populations was set at 5% based on a review of published stray rates for Chum Salmon in British Columbia (McElhany et al. 2000).

Chum Salmon abundance in Southern BC is largely estimated from visual surveys (by foot and air), which typically produce relatively imprecise estimates of abundances. In the observation sub-model, we assumed the standard deviation in observations 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 et al. 1982, Szerlong and Rundio 2008). 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 CU-specific catch estimates, 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) using catch and recruitment data from two DFO Fishery Statistical Areas of Chum Salmon on the west coast of Vancouver Island, BC (Dobson et al. 2009). Because outcome uncertainties likely vary among stocks and management approaches, we also considered an upper value of 0.5 in a sensitivity analysis. Here we assumed magnitude of outcome uncertainties did not vary with stock size. Alternative assumptions such as larger uncertainties at small stocks sizes could be considered in future analyses.

3.1.3 Sensitivity Analyses

3.1.3.1   Univariate, Global, and Bivariate Sensitivity Analyses of Benchmarks

To assess the strength and direction of effects of input parameters on benchmark performance (measured as deviations between estimated benchmarks and "true" benchmarks), we performed a sensitivity analysis where each input parameter was varied individually while all others were held constant at "base-case" values (Table 5). However, this analysis did not assess sensitivity to interactions among input variables. To further consider interactions among all input variables, we performed a global sensitivity analysis using the Morris method (Morris et al. 2014). Similar to univariate analyses, the Morris method varies each input parameter one at a time, but in contrast to univariate analyses, this is done with different combinations of other variables. The sensitivity of benchmark performance to uncertainty in each parameter given uncertainty in other parameters is measured by a composite metric of the overall influence of a parameter on the output, labeled the mean 'elemental effect' (Morris et al. 2014). 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 the two parameters that had relatively large effects on performance, productivity and harvest rates, a bivariate (two-way) sensitivity analysis was performed to assess their combined effect on benchmark performance. In addition, given estimates of productivities for CUs from Section 2.3.1 and observed harvest rates (Appendix A), we evaluated CU-specific performances. We focused univariate and global sensitivity analyses on lower benchmarks (25th percentile and S_{gen}), but also considered sensitivities of upper benchmarks (75th percentile and 80% S_{MSY}) in the bivariate sensitivity analysis.

Based on results of the bivariate sensitivity analyses showing that the percentile-based benchmark, S_{25th} was lower than "true" underlying data-rich stock-recruitment benchmark, S_{gen} in some situations, we evaluated additional percentile-based lower benchmarks based on 30th to 50th percentiles in increments of 5%. We also evaluated the performance of additional upper benchmarks at the 50th-70th percentile of observed spawner abundances (increments of 5%) given results showing that the S_{75th} percentile benchmark far exceeded the "true" upper benchmark, 80% S_{MSY}, in some situations. Consequently, the 50th percentile of spawner abundances was evaluated as both an upper and lower benchmark.

We further considered bivariate sensitivity on two components of observation errors, magnitude of errors in spawner abundances and biases in spawner abundances, due to for example, partial sampling.

3.1.3.2 *Sensitivity Analyses on Stock-Recruitment Parameters and Data Contrast*

Performance of data-rich benchmarks depends in part on the accuracy and precision of estimated stock-recruitment parameters. We further evaluated mean percent error between "true" and estimated stock-recruitment parameters (instead of benchmarks) along gradients in "true" productivity and harvest rates to assess how errors in these parameters influence performance of the data-rich benchmarks.

Performance of both data-rich and data-limited benchmarks may also depend on the magnitude of contrast in historical spawners time-series as described by Clark et al. (2014). We evaluated the impact of variability in harvest rates and productivity on data contrast in our simulation model, to assess covariation in those variables and the value of considering data contrast when productivity and harvest rate information is not available.

3.2 RESULTS

Time-series of spawners and recruitment for an example Monte Carlo trial from the simulation are presented in Figure 10a. Percentile benchmarks tended to vary annually (as shown by red and green dashed lines for S_{25th} and S_{75th}, respectively, Fig. 10a), whereas stock-recruitment benchmarks were more consistent over time (red and green solid lines, Fig. 10a). To estimate stock-recruitment based benchmarks, Ricker curves were fit annually using only data up until the year being assessed (Fig. 10b for the last year of the simulation).
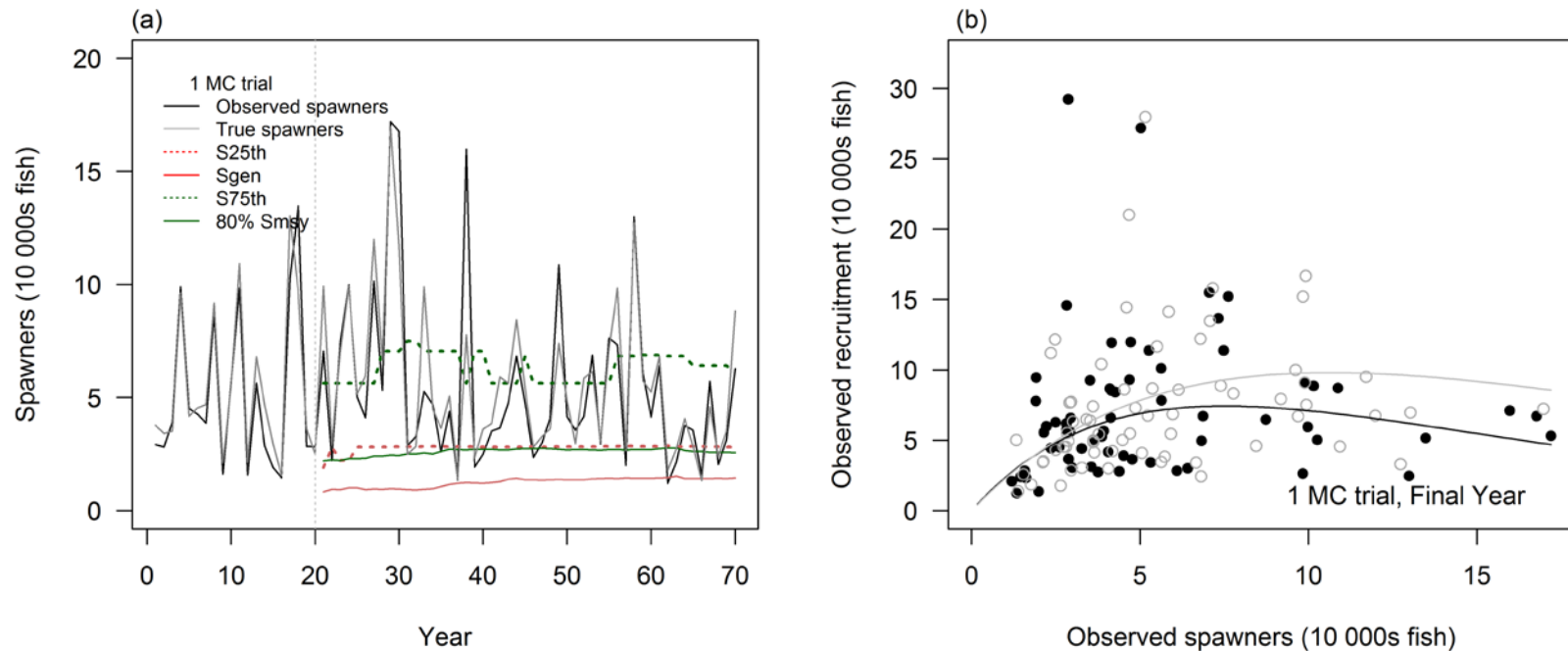
Figure 10. (a) Time-series of observed spawner abundances (black line) and "true" spawner abundances (grey line) and benchmarks for one Monte Carlo trial for a hypothetical CU. Vertical dashed line indicates the end of the 20-year initialization period of the simulation. Benchmarks are estimated annually based on all data up until that year: 80% S_{MSY} (upper benchmark, green solid line), S_{gen} (lower benchmark, red solid line), S_{75th} percentile benchmark (green dotted line), and the S_{25th} percentile (red dotted line). (b) Observed spawner and recruitment data (solid black dots) and "true" data (grey hollow dots) for the final year of the Monte Carlo trial depicted in panel (a). The "true" underlying stock-recruitment relationship is shown with the grey curve. The estimated curve based on observed data is shown in black. (c) Mean percent error between estimated and "true" benchmark averaged over all years and 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% of S_{MSY} (upper benchmark), on next page.
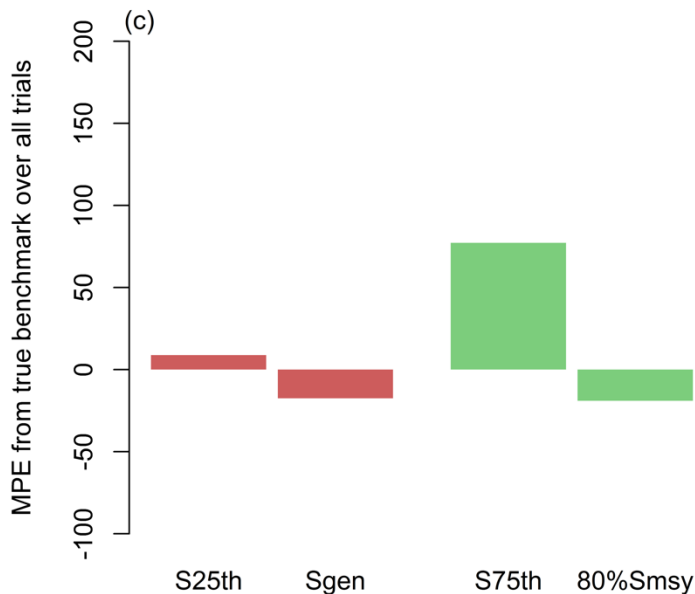
Figure 10 continued.

In this hypothetical example, percentile benchmarks, S_{25th} and S_{75th}, tended to be higher than the "true" benchmark values (MPEs were positive), whereas the estimated stock-recruitment benchmarks, S_{gen} and 80% of S_{MSY}, tended be negatively biased (MPEs were negative) (Figure 10c). Estimates of stock-recruitment benchmarks differed from the "true" values because they were based on observed data (black line in Figure 10b, solid dots in 10b) rather than "true" data (grey line in Figure 10a, hollow dots in Figure 10b). The assessed stock-recruitment model (black curve, Figure 10b) differed from the "true" underlying model (grey curve, Figure 10b) due to, for example, errors in spawner abundances and recruitment, and time-series biases (Walters and Martell 2004).

3.2.1 Univariate and Global Sensitivity Analyses

In univariate sensitivity analyses, performances of lower benchmarks (both estimates of S_{gen} and S_{25th}) were more sensitive to uncertainty in productivity than to other input parameters (Figure 11). Low productivity was associated with estimates of benchmarks that fell below "true" values (Fig. 11, leftmost black bar), resulting in estimates status that were overly optimistic compared with the "true" status. In contrast, high productivity was associated with estimates of benchmarks that were above the "true" values (Fig. 11, leftmost white bar), resulting in more depleted estimates of status compared with the "true" status. For the lower percentile benchmark, S_{25th}, harvest rates had a strong effect on benchmark performance, and the remaining input parameters had relatively weak effects on performance. For the lower benchmark, S_{gen}, the Ricker autocorrelation coefficient had moderate impacts on performance and the remaining input parameters had relatively weak effects on performance.
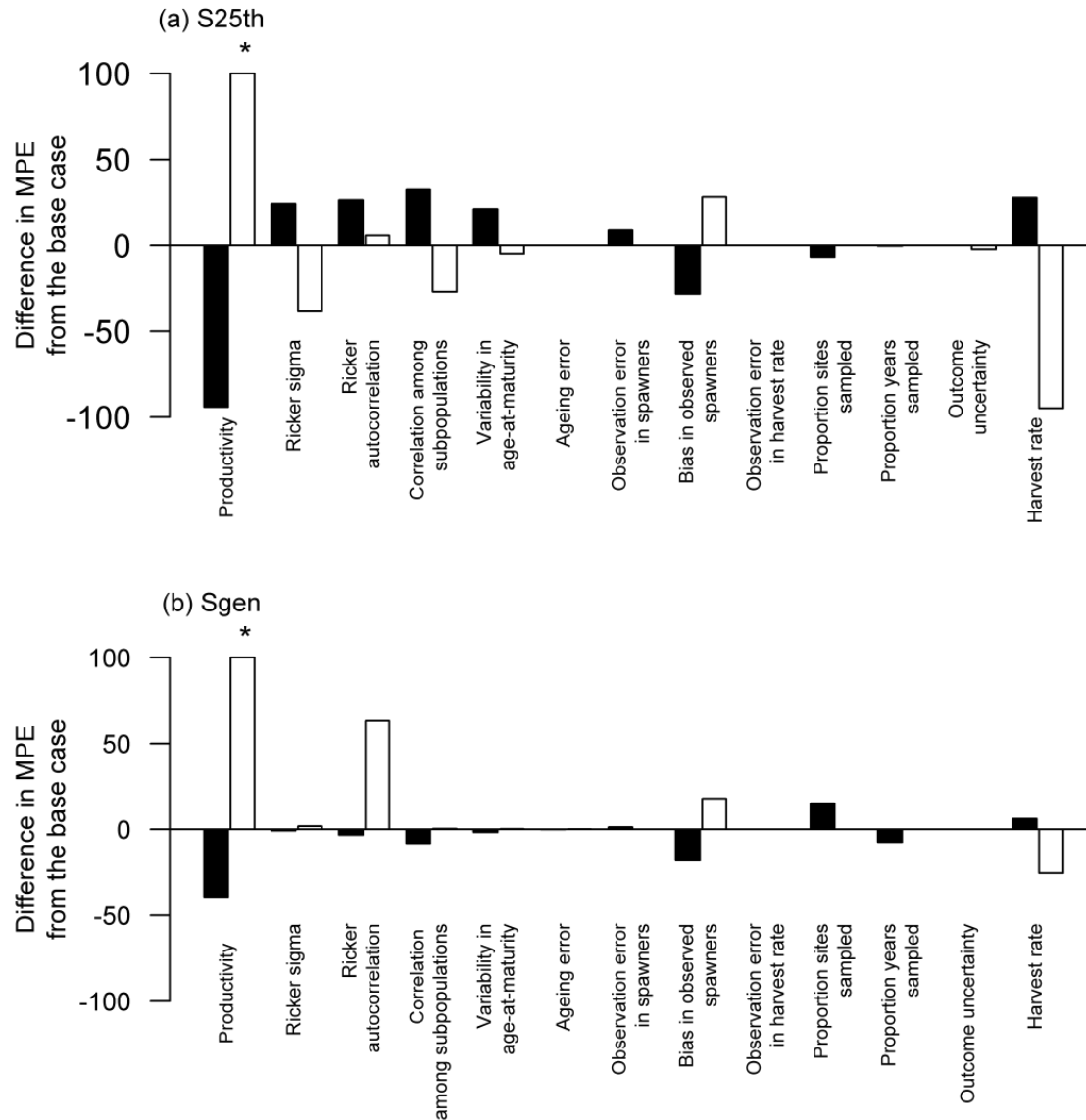
Figure 11. Difference in the mean percent error, MPE, between sensitivity analyses listed on the x-axis and the base case scenario for percentile lower benchmark, S_{25th} (a) and estimated stock-recruit lower benchmark, S_{gen}, (b). Positive values indicate sensitivity analyses where MPE increased under that change in input parameter; negative values indicate analyses where the MPE declined under that change in input parameter. Black bars are analyses where the input parameter was increased relative to the base case (see Table 5); white bars are analyses where the input parameter was reduced relative to the base case. Asterisks denote values higher than the limit of the y-axis: 132% (a) and 586% (b).

Results from global sensitivity analyses were similar to those of the univariate analyses. The effects on performance were greatest for productivity for both S_{25th} and S_{gen} benchmarks (x-axis of mean elemental effect on Fig. 12). Harvest rates were secondarily important for the S_{25th} benchmark (Fig. 12a). Parameters that ranked high on the standard deviation in the elemental effect (y-axis of Fig.12b, e.g., observation errors in spawner) were influential for benchmark performance only in combination with other input parameters. In other words, S_{gen} performance was only sensitive to variability in observation errors in spawner abundances under certain combinations of other variables (Fig. 12b, top).
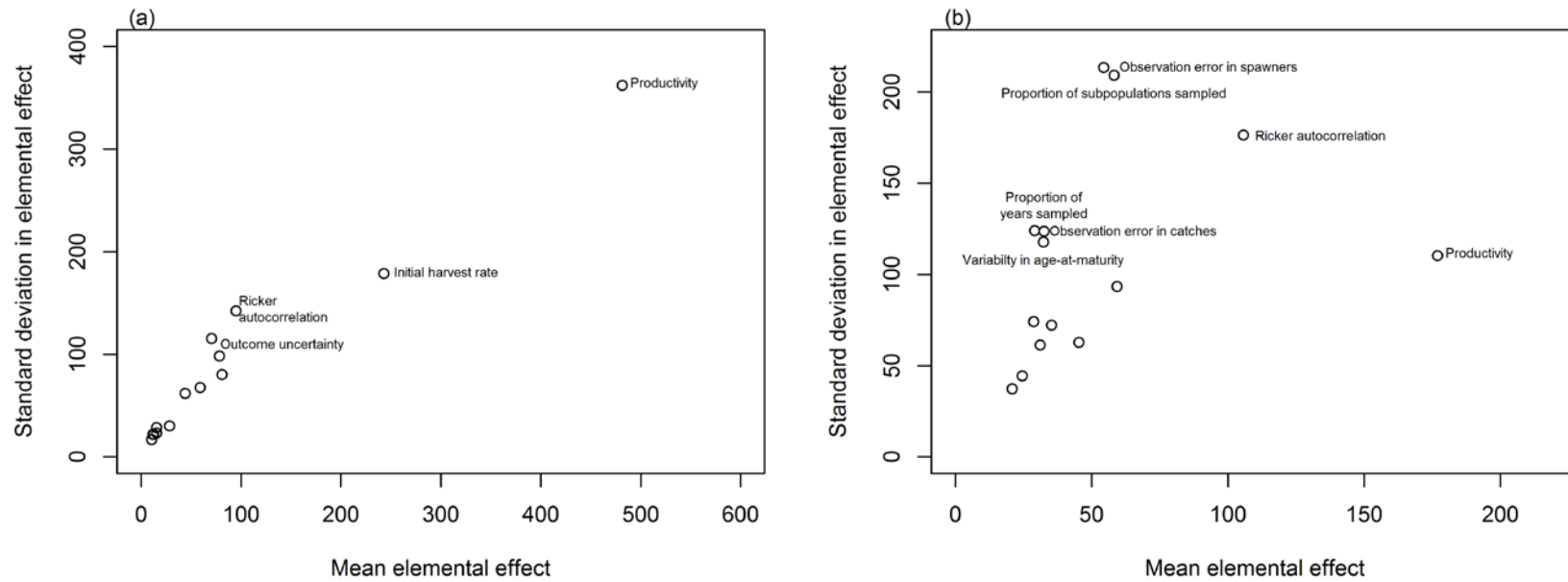
Figure 12. Sensitivity indices of the effects of individual variables (mean elemental effect, x-axis) and interactions among variables (standard deviation in elemental effects, y-axis). Indices were derived from the Morris method, a global sensitivity analyses for the mean percent error of estimated lower benchmarks, S_{25th} (a), and S_{gen} (b), from "true" benchmarks. Input variables with values >100 on either axis are labelled.

3.2.2 Bivariate Sensitivity Analyses: Lower Benchmark

We further explored the effects of variability in productivity and harvest rates on benchmark MPEs in bivariate (two-way) sensitivity analyses. At moderate to high productivity and low initial harvest rates, estimates of both S_{25th} and S_{gen} benchmarks were equal to or higher than the "true" S_{gen} lower benchmark and can be considered precautionary from a conservation perspective (Figure 13, top left portion of panels). However, at low productivity and high harvest rates, neither benchmark was above the "true" benchmark (Figure 13 bottom right portion of panels), and this was true for S_{gen} even at low harvest rates (Figure 13, bottom left corner).
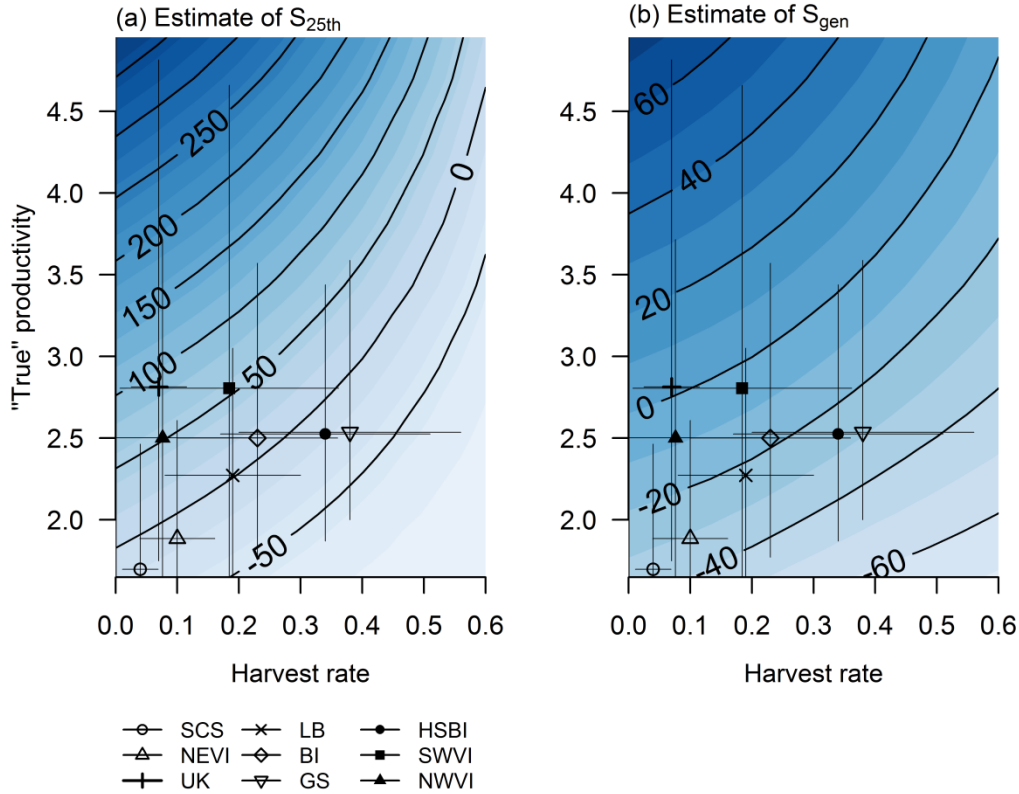
Figure 13. Mean percent error, MPE, of the estimated lower benchmark, S_{25th} (a), and S_{gen} (b) from the "true" lower benchmark ("true" S_{gen}) depicted on isopleths, along a gradient in harvest rates and "true" productivities. MPEs were derived from a simulation model of a hypothetical Chum Salmon CU. Symbols indicate CU-specific harvest rates and productivities. 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 Northeast Vancouver Island, UK is Upper Knight, LB is Loughborough, GS is Georgia Strait, and HSBI is Howe Sound/Burrard Inlet, NWVI is Northwest Vancouver Island, and SWVI is Southwest Vancouver Island.

Given CU-specific estimates of productivities and average harvest rates, we identified variability in benchmark performance among CU (Figure 13a). The S_{25th} benchmarks were near "true" S_{gen} (near the zero contour lines) for 1 CU, Loughborough, higher than "true" S_{gen} for 4 CUs, Northwest Vancouver Island, Southwest Vancouver Island, Bute Inlet, and Upper Knight, and below "true" S_{gen} for 4 CUs, Southern Coastal Streams, Northeast Vancouver Island, Georgia

Strait, and Howe Sound to Burrard Inlet. Estimates of S_{gen} fell below the "true" values for all CUs except Upper Knight (Figure 13b).

We also evaluated CU-specific performance of alternative lower percentile benchmarks ranging from the 30th – 50th percentiles in increments of 5%. We found that the 50th percentile benchmark tended to be higher than the "true" estimate S_{gen} for all CUs, but uncertainty intervals for several CUs crossed the zero contour line (Figure 14a). Plots of the performance of remaining increments of percentile-based benchmarks are provided in Appendix G.
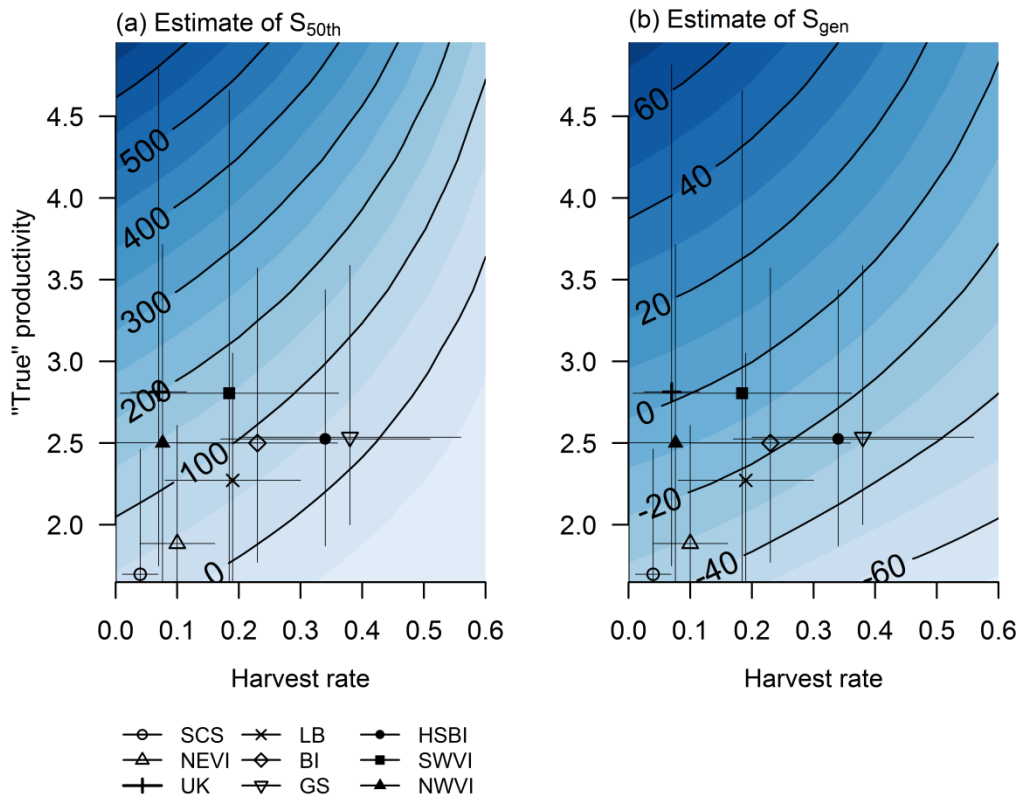
Figure 14. Mean percent error, MPE, of the estimated lower benchmark based on the 50th percentile of observed abundances, S_{50th}, (a), and S_{gen} (b), from the "true" lower benchmark, S_{gen}, depicted on isopleths, along a gradient in harvest rates and "true" productivities. MPEs were derived from a simulation model of a hypothetical Chum Salmon CU. See the caption for Fig. 13 for an explanation of symbols, lines, and abbreviations. Panel (b) is the same as in Fig. 13b, but is shown to facilitate comparison with alternative percentile-based benchmarks.

To compare performance of percentile benchmarks across various percentiles (25th, 30th, 35th, 40th, 45th, and 50th), we further categorized productivities and harvest rates into 3 groups each (productivity: ≤2.5, >2.5 and ≤4, and >4 recruits/spawner; harvest rates ≤20%, >20% and ≤40%, and >40% and ≤60%) and aggregated performance of percentile benchmarks within groups. Figure 15 depicts the aggregate performance of various lower percentile benchmarks represented by box and whiskers, over all combinations of productivities (rows) and harvest rates (columns). White boxes indicate percentile benchmarks where the lower bound of the

confidence range (bottom whisker) is greater than zero, but this difference is minimized. These reflect percentile benchmarks that are relatively close to the "true" S_{gen}, but have a high probability of being above it. Where the lower bound of the distributions of all percentile benchmarks fall below zero (i.e., are lower than "true" S_{gen}, Fig. 15f, h and i), further evaluation is required to assess if alternative percentile benchmarks can be used without risks of biased (particularly overly optimistic) status assessments. Alternative groupings for productivity were considered (≤2, >2 and ≤4, and >4 recruits/spawner) and the results were similar, with the exception of the combined moderate productivity and harvest rate category (Appendix G, Fig. G5e).

Figure 15 Distributions of mean percent error, MPE, of various percentile lower benchmarks (labelled on x-axis) from "true" S_{gen} along a gradient in harvest rates (columns, labelled at top) and productivities (rows labeled on right), derived from a simulation model of a hypothetical Chum Salmon CU. Distributions for each benchmark are derived from simulated outputs in Fig. 13, Fig. 14 and Appendix G, aggregated within productivity and harvest rate groupings. White boxes indicate percentile benchmarks where the lower confidence bound ≥0 and this difference is minimized. "Further evaluation required" signifies cases where all percentile benchmarks had lower confidence bounds <0.

Note, by aggregating results within productivity and harvest rate categories, fine-scale variability in performance along gradients in those variables is lost. For example, Fig. 12 shows that for

the CU, Northwest Vancouver Island (NWVI), the S_{25th} benchmark is greater than the "true" S_{gen} benchmark and hence would provide a precautionary estimate of status that is equal to or lower than the "true" status. However, this CU falls within the low productivity and low harvest rates group (though at the extreme edge, productivity = 2.5, harvest rate =8%), where S_{50th} is suggested as a lower benchmark, not S_{25th} (Fig. 15g).

When information on productivity is not available, performance can be compiled across the three productivity zones (Fig. 16). In this case, further evaluation is warranted when harvest rates are ≥20%.

Figure 16. Distributions of mean percent error, MPE, of various percentile lower benchmarks (labelled on x-axis) from "true" S_{gen} along a gradient in harvest rates (columns, labelled at top), compiled over the range of productivities considered here, 1.5-5 recruits/spawner White boxes indicate percentile benchmarks where the lower confidence bound ≥0 and this difference is minimized. "Further evaluation required" signifies cases where all percentile benchmarks had lower confidence bounds <0.

3.2.3 Bivariate Sensitivity Analyses: Upper Benchmark

The S_{75th} upper benchmark was often considerably higher than the "true" 80% of S_{MSY} benchmark, especially at high productivity and low harvest rates (Figure 17a). However, at low productivity and high harvest rates, S_{75th} tended to underestimate the "true" upper benchmark. In addition, estimates of 80% of S_{MSY} tended to be below the "true" value except at high productivity (Figure 17b), but the magnitude of these errors was relatively small.

For the CUs examined here, the S_{75th} benchmarks were higher than the "true" 80% of S_{MSY}, and sometimes exceeding them by >150% (symbols on Figure 17a). However, for 2 CUs, the uncertainty bounds of the S_{75th} benchmark crossed the zero contour line. The estimates of 80% of S_{MSY} were below the true benchmark values for all CUs (Figure 17b).
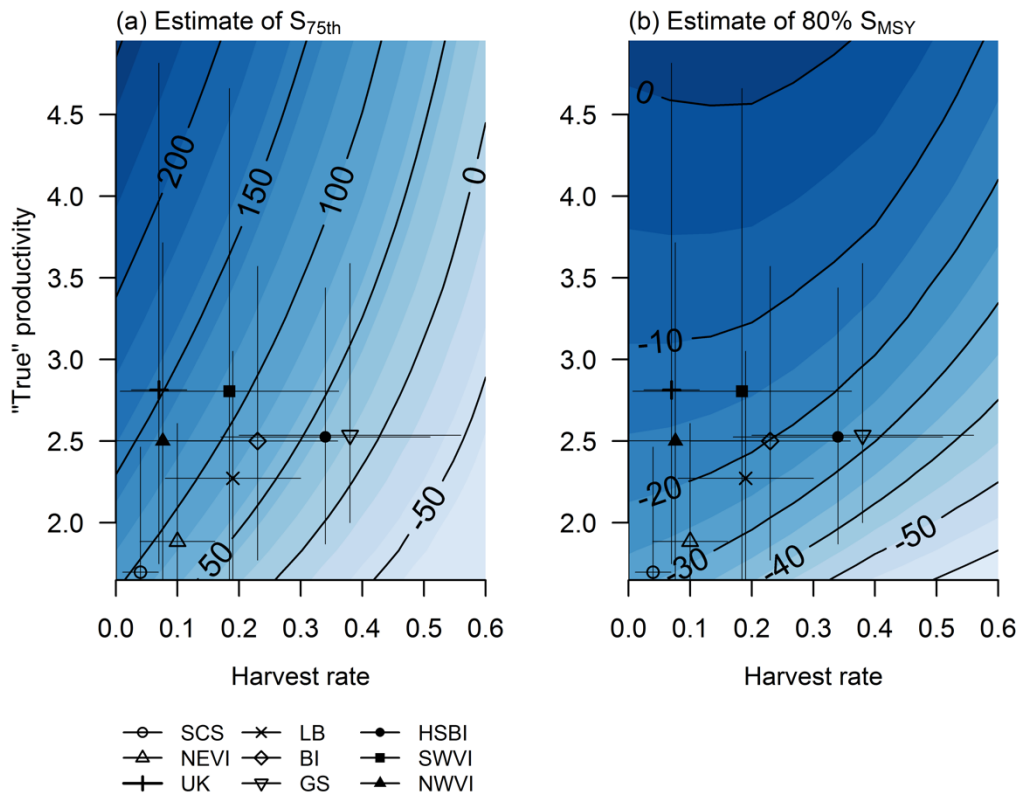
Figure 17. Mean percent error, MPE, of the estimated upper benchmarks, S_{75th} (a), and 80% of S_{MSY} (b), from the "true" 80% of S_{MSY} depicted on isopleths, along a gradient in harvest rates and "true" productivities. MPEs were derived from a simulation model of a hypothetical Chum Salmon CU. See the caption for Fig. 13 for an explanation of symbols, lines, and abbreviations.

Given that the S_{75th} benchmark significantly overestimated the "true" upper benchmark in some situations, we also investigated performance of alternative upper benchmarks, S_{50th}-S_{70th}, in increments of 5%. The S_{50th} benchmark is more closely aligned with benchmarks on long-term trends in spawner abundances currently used for WSP status assessment than S_{75th} (Grant et al. 2011). The long-term trend metric under the WSP captures the ratio of the current spawner abundances (geometric mean over the current generation) to the historical geometric mean, with lower and upper benchmarks at 0.5 and 0.75, corresponding to values below the historical mean and likely less than then median, S_{50th}. Metrics on long-term trends in spawner abundances under the WSP are not used to assess status on their own, but are combined with additional metrics to provide integrated status assessments (e.g., Grant and Pestal 2013).

We found that the S_{50th} upper benchmark tended to be closer in value to the "true" upper benchmark than S_{75th}, but that the S_{50th} underestimated the "true" upper benchmark when harvest rates were high (Figure 18). For 3 CUs, Northwest Vancouver Island, Southwest Vancouver Island, and Upper Knight, S_{50th} benchmarks were >20% larger than "true" upper benchmarks. For the remaining CUs, S_{50th} benchmarks were near or below "true" upper benchmarks.
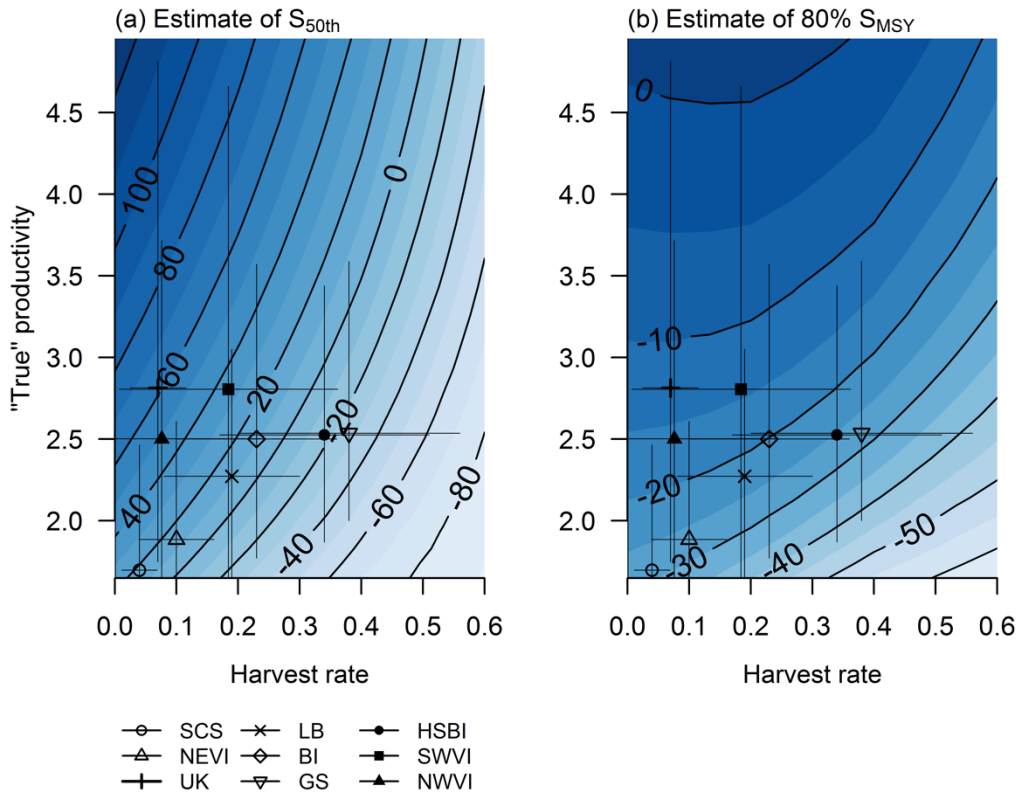
Figure 18. Mean percent error, MPE, of the estimated upper benchmarks, S$_{50th}$ (a), and 80% of S$_{MSY}$ (b), from the "true" 80% of S$_{MSY,}$ depicted on isopleths, along a gradient in harvest rates and "true" productivities. MPEs were derived from a simulation model of a hypothetical Chum Salmon CU. See the caption for Fig. 13 for an explanation of symbols, lines, and abbreviations.

Similar as for the lower benchmark, to compare performance of percentile upper benchmarks across various percentiles (50th, 55th, 60th, 65th, 70th, and 75th), we aggregated performance of percentile benchmarks over combinations of productivities and harvest rates, according to the groups described in Section 3.2.2. However, in contrast to the lower benchmarks, for upper benchmark performance, white boxes indicate percentile benchmarks where the median performance (solid horizontal lines at the mid-point of the boxes) is greater than zero and this difference is minimized. This choice reflects percentile benchmarks that are relatively close to the "true" 80% of S$_{MSY}$, and have at least a 50% probability of being above it. In contrast to the lower benchmark which is set at a level to avoid listings of endangerment by COSEWIC, the upper benchmark is set at a level to achieve maximum annual catches on an average annual basis, allowing for years when abundances drop below this level. The choice of avoiding negative biases for lower benchmark but not upper benchmark was subjective, but reasonable in light of the benchmark definitions under the WSP. Different assumptions could be used to identify specific percentiles within the range identified here, based on results shown on Figs 15 and 18. Highlighted percentiles are shown for illustrative purposes, and because they are aligned with the intent of the WSP. However, alternatives could be considered. "Further evaluation required" indicates those categories where either lower or upper benchmarks fall

below these values. Alternative groupings for productivity provided similar results (Appendix G, Fig. G6).
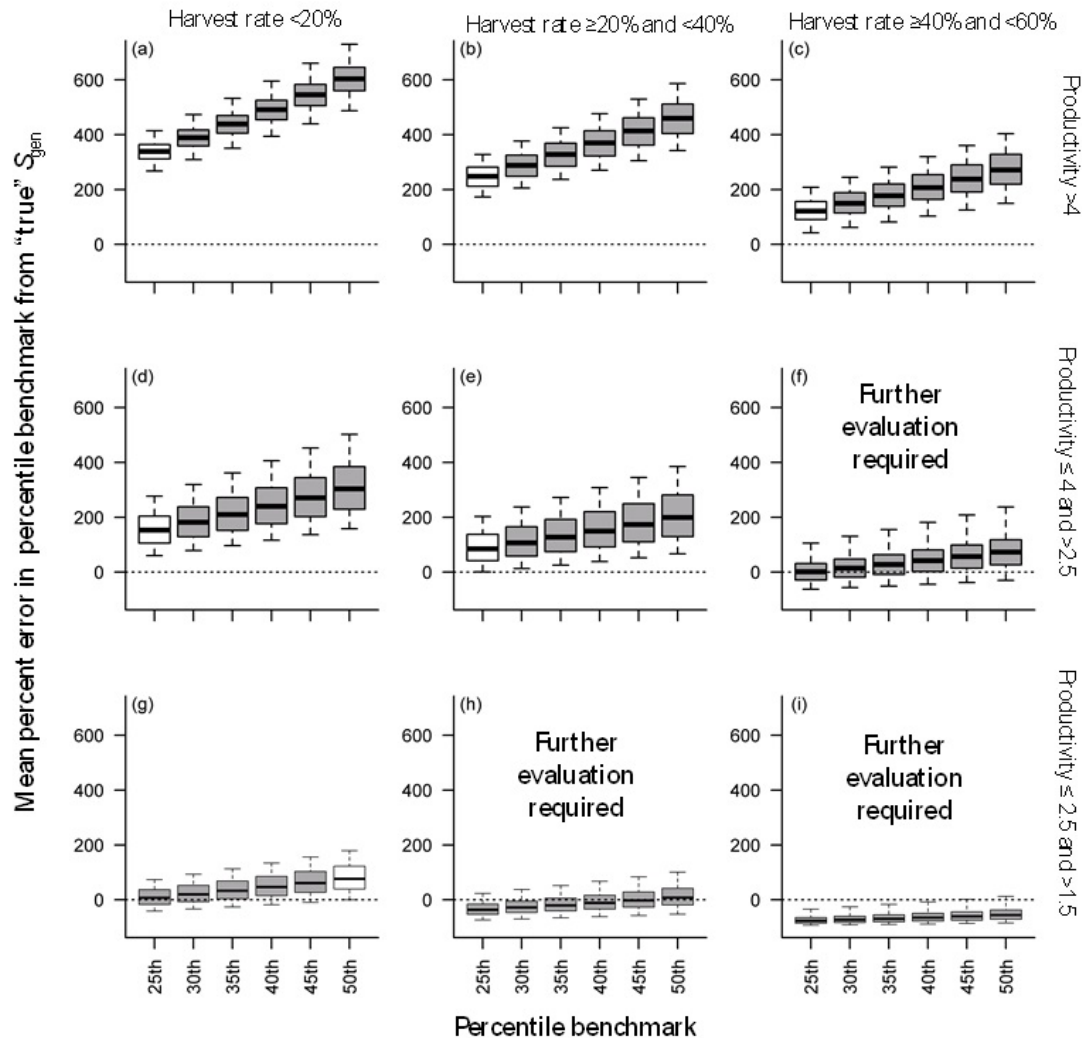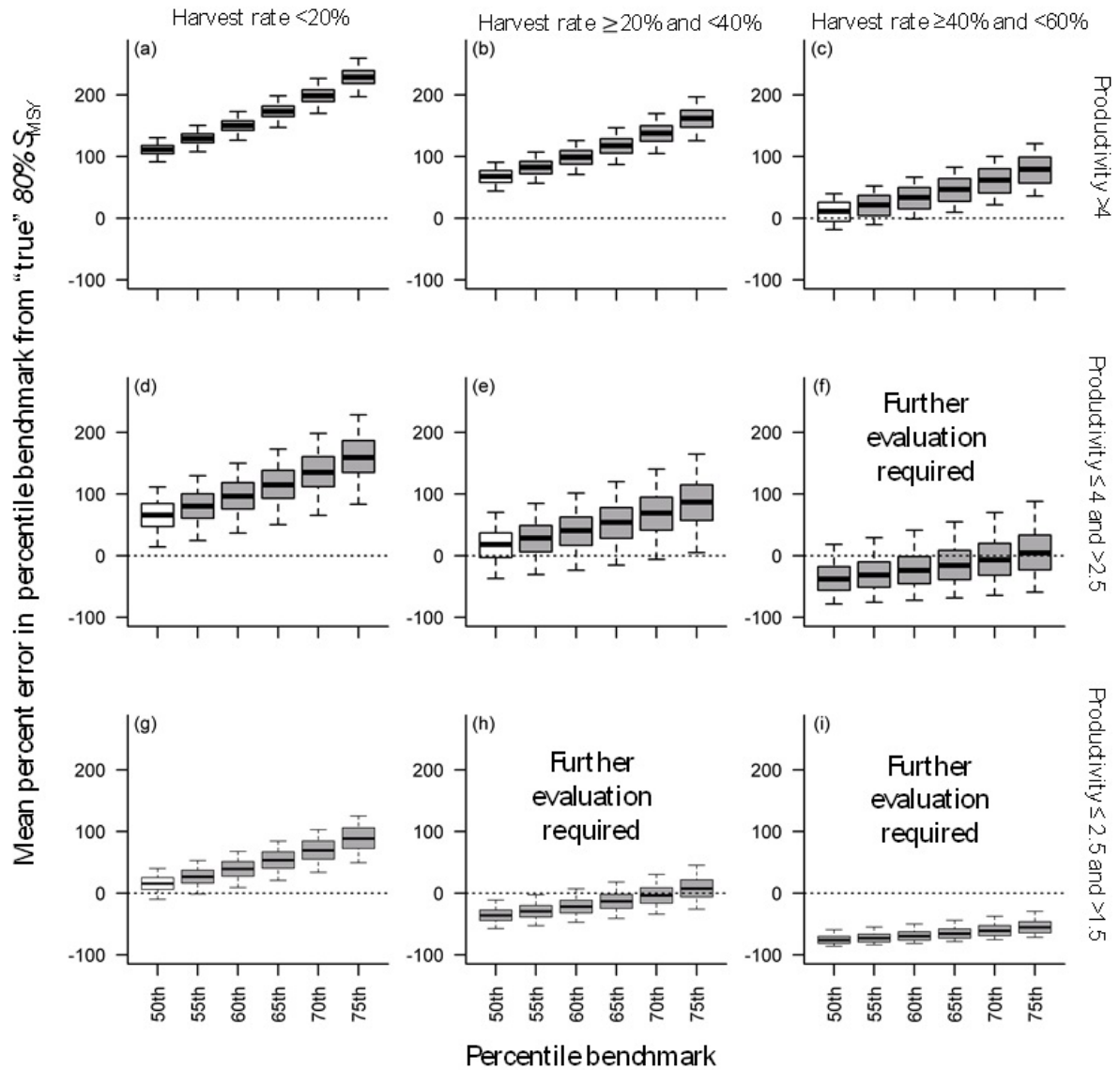

Figure 19 Distributions of mean percent error, MPE, of various percentile lower benchmarks (labelled on x-axis) from "true" S_{gen} along a gradient in harvest rates (columns, labelled at top) and productivities (rows labeled on right), derived from a simulation model of a hypothetical Chum Salmon CU. Distributions for each benchmark are derived from simulated outputs, aggregated within productivity and harvest rate groupings. White boxes indicate percentile benchmarks where the median performance value ≥0 and this difference is minimized. "Further evaluation required" indicates cases where lower benchmarks could not be selected (Fig. 15).

When information on productivity is not available, performance can be compiled across the 3 productivity zones (Fig. 20). Although the median performance of S_{50th} is >0 at moderate harvest rates (≥20% and <40%, Fig. 20b) and the median performance of S_{75th} is >0 at high harvest rates (≥40% and <60%, Fig. 20c), further evaluation is suggested for these categories because lower benchmarks are not identified in those cases (Fig. 16b and c)
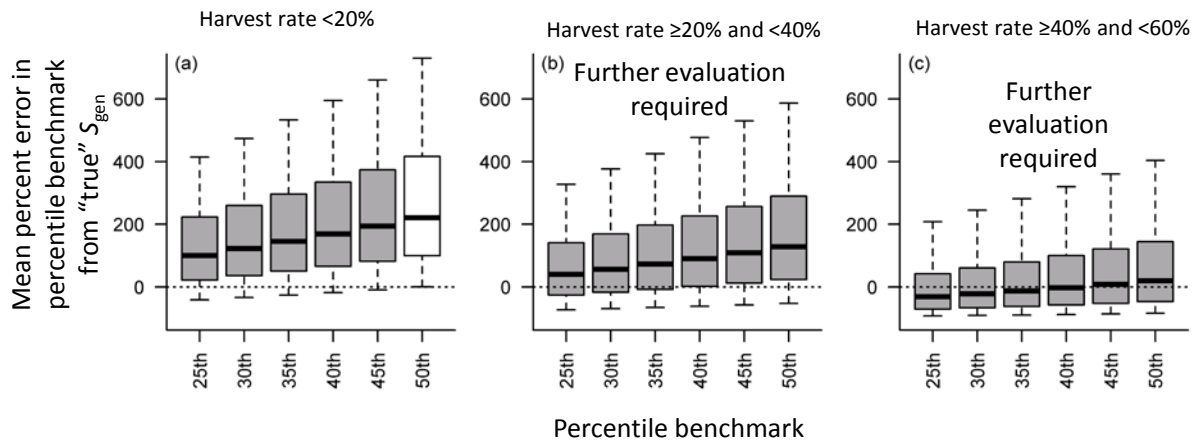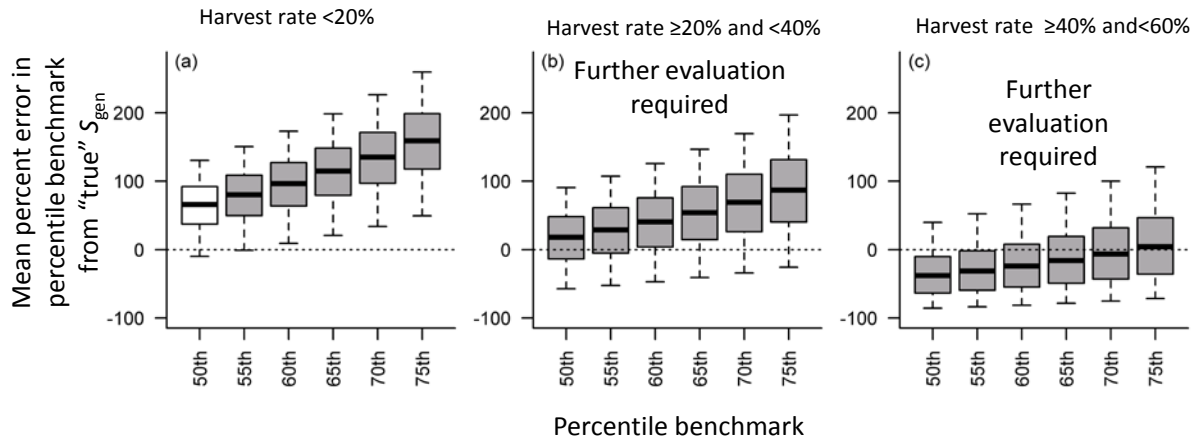
Figure 20. Distributions of mean percent error, MPE, of various percentile upper benchmarks (labelled on x-axis) from "true" 80% S_{MSY} along a gradient in harvest rates (columns, labelled at top), compiled over the range of productivities considered here, 1.5-5 recruits/spawner White boxes indicate percentile benchmarks where the lower confidence bound ≥0 and this difference is minimized. "Further evaluation required" signifies cases where all percentile benchmarks had lower confidence bounds <0.

3.2.4 Bivariate Sensitivity Analyses: Ricker Parameters and Data Contrast

To investigate the impacts of biases in underlying stock-recruitment parameters on the performance of data-rich benchmarks, S_{gen} and 80% of S_{MSY}, we further evaluated deviations of estimated Ricker parameters from "true" values along gradients in "true" productivity and harvest rates. We found similar patterns in deviations in estimated Ricker productivity parameter from the "true" values, as for S_{gen} and 80% S_{MSY} deviations, though the direction of effects differed (Figure 21). At low productivity and high harvest rates, the estimated productivity (Ricker α) tended to be higher and estimated capacity (S_{max}) and equilibrium abundances (S_{eq}) tended to be lower than the "true" values (Figure 21, bottom right corner of all panels). The opposite occurred at high productivity and low harvest rates.
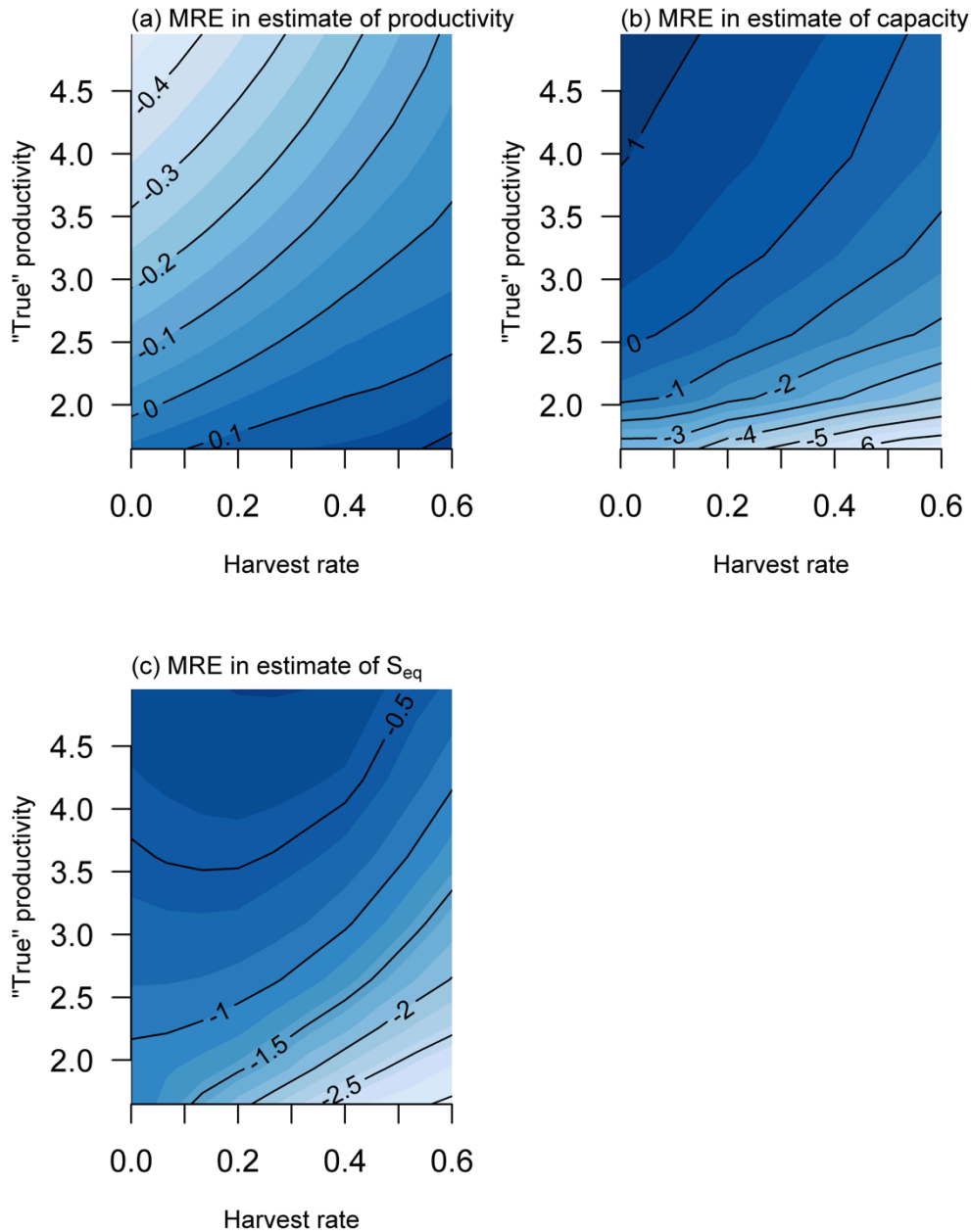
Figure 21. Mean raw error, MRE, of estimated productivity (a), estimated carrying capacity, S_{max} (b), and spawner abundances at equilibrium, S_{eq} (c), depicted on isopleths, along gradients in "true" productivity and harvest rates. MREs were derived from a simulation model of a hypothetical Chum Salmon CU.

This overestimation of productivity and underestimation of capacity (especially at low productivity and high harvest rates) results in estimates of S_{gen} and S_{MSY} benchmarks that fall below "true" values, as documented for an example stock-recruitment curve (Figure 22), and by Holt and Folkes (2015). Although the magnitude of the biases will vary among CUs, the direction of biases on stock-recruitment parameters and benchmarks will be consistent.
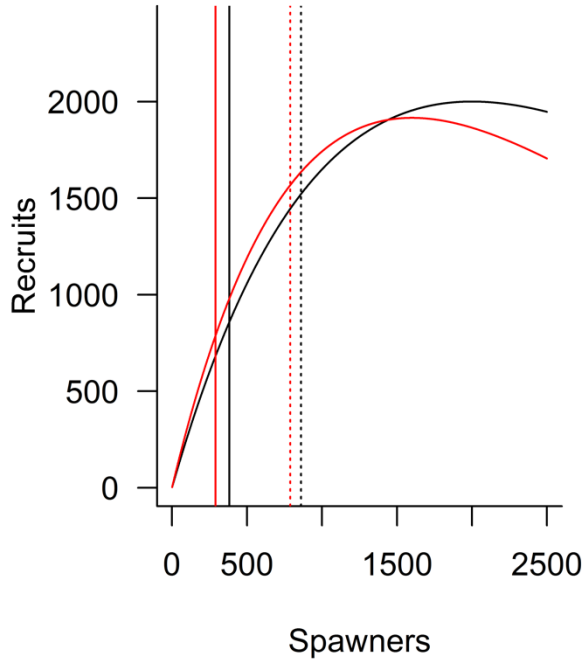
Figure 22. Stock-recruitment curves for a hypothetical CU under a base case of moderate productivity (=2.7) and capacity of 2000 (black curve), and a scenario where productivity is overestimated by 20% and capacity is underestimated by 20%, as is typical of time-series biases (red curve). Solid vertical lines represent S_{gen} benchmarks for the base case (black) and biased parameter estimates (red); dotted lines represent S_{MSY} for the base case (black) and biased parameters (red).

Our model assumed spawner abundances at equilibrium, S_{eq}, remained constant as productivity varied in sensitivity analyses based on the Ricker stock-recruit model (Appendix F, Eqn. F15). We used this assumption to be consistent with previous models that evaluated benchmarks under the WSP (Holt 2009, Holt and Bradford 2011). In sensitivity analyses, we considered an alternate assumption where S_{max} remained constant, but S_{eq} declined as productivity declined, which represents a variation of the Ricker model commonly used in the literature and applied in the retrospective analyses here (Eqn. 1). Under this assumption, we found similar patterns in the results (within ~10% mean percent error in benchmarks). Sensitivity analyses on alternative formulations of the Ricker model were also performed by Holt and Folkes (2015) for data-rich benchmarks. Similar to our results, they found that benchmark performance was robust to the formulation of the Ricker model.

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 (Figure 23).
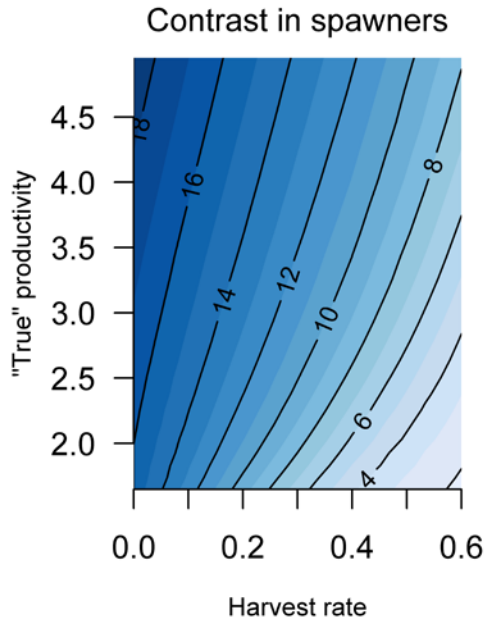
Figure 23. Contrast in spawner escapement data (maximum escapement/minimum escapement) over gradients in "true" productivity and harvest rates derived from a simulation model of a hypothetical Chum Salmon CU.

3.2.5 Bivariate Sensitivity Analyses: Observation Errors

The effects of the magnitude of observation errors in spawner abundances were small (+/- 5%) compared with the effects of biases in spawner abundances (+/- 35%) (Figure 24), but both were smaller than the effects of productivity and harvest rates (Figure 24 compared with Figure 13).
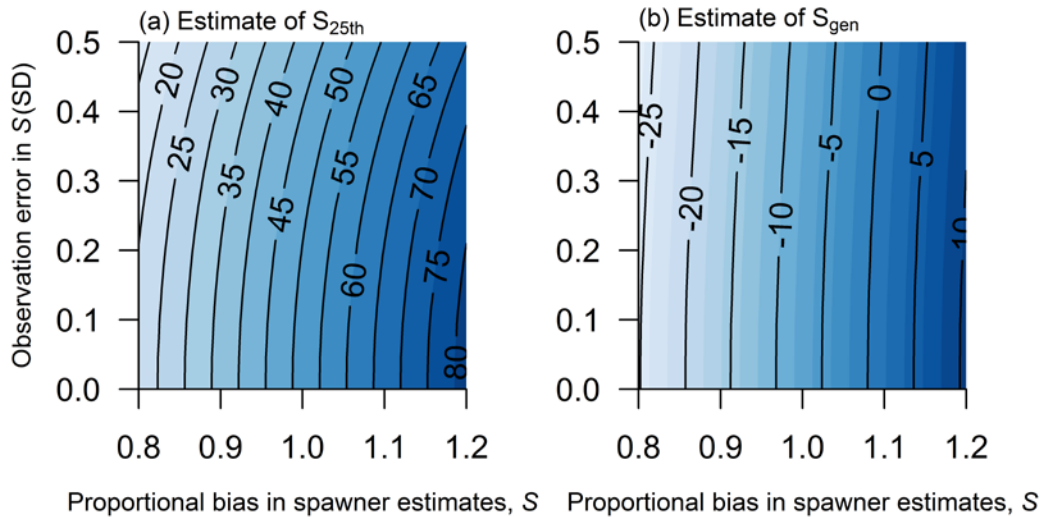
Figure 24. Mean percent error, MPE, of the estimated lower benchmark, S_{25th} (a), and S_{gen} (b), from the "true" S_{gen} value, depicted on isopleths, along a gradient in bias in estimated spawner abundance and observation errors in spawner abundances (SD). MPEs were derived from a simulation model of a hypothetical Chum Salmon CU.

3.3 DISCUSSION

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. For highly productive CUs with low harvest pressure, time-series were dominated by high abundances pushing the S_{25th} benchmark upwards above the "true" lower benchmark. For example, for 5 CUs (Northwest Vancouver Island, Southwest Vancouver Island, Loughborough, Bute Inlet, and Upper Knight), the S_{25th} lower benchmarks were equal to or higher than "true" S_{gen} benchmarks, providing similar or more precautionary estimates of status.

In contrast, S_{25th} benchmarks tended to be below the "true" lower benchmarks when harvest on unproductive CUs was high. When CUs were depleted, the time-series of observed abundances were dominated by low values, ratcheting the S_{25th} benchmark downward over time. We found that an alternative benchmark, S_{50th}, derived from median of the spawner time series, was higher than "true" lower benchmarks for the remaining 4 CUs (Southern Coastal Streams, Northeast Vancouver Island, Howe Sound-Burrard Inlet, and Georgia Strait).

However, under various combinations of moderate-high harvest rates and low-moderate productivities, even the lower benchmark, S_{50th}, tended to be below the "true" S_{gen} benchmark (Figure 14, bottom right corner). This occurred when harvest rates were approximately ≥40% and productivities were approximately <4 recruits/spawner, or harvest rates were between 20-40% and productivities were low (≤2.5 recruits/spawner). Under these scenarios, percentile-based benchmarks may overestimate status, possibly generating relatively healthy status assessments when conservation concerns may exist. Further research into alternative percentiles or other types of benchmarks (e.g., derived from habitat capacity) are warranted in this case.

Estimates of S_{gen} tended to be negatively biased when productivity was low due to time-series biases on stock-recruitment parameters. Time-series biases occur when spawner abundances, the independent variable in stock-recruitment relationship depends on the recruitment in the

previous generation, the dependent variable. These biases are well documented for salmon populations (Walters and Martell 2004). The lack of independence between spawners and recruitment results in overestimates of productivity and underestimates in capacity (as documented for Skeena River salmon in BC, Korman and English 2013), and these effects are accentuated for unproductive stocks (Korman et al. 1995). These parameter biases resulted in underestimates of S_{gen} and 80% of S_{MSY}. Although state-space versions of stock-recruitment models 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 described in the retrospective analyses (Section 2) or models with environmental covariates have been suggested as a method to reduce these biases (Korman and English 2013). In addition, alternative formulations of the Ricker model with management-relevant parameters (maximum sustainable catch and harvest rate) have been found to reduce these biases (Schnute and Kronlund 1996). However, a thorough evaluation of these methods under different scenarios of productivity and data quality is lacking.

Our findings also suggest that estimates of data-rich benchmarks may not provide accurate measures of status when estimated from uncertain data, especially if time-series biases occur in the assessment procedure. A thorough investigation of data-rich benchmarks was outside the scope of this study. Further work investigating alternative stock-recruitment models, which may mitigate the impacts of time-series biases is warranted, especially if combined with a simulation evaluation of associated management procedures.

To inform assessments for Alaskan Pacific salmon, Clark et al. (2014) identified sustainable escapement goals based on percentiles of observed spawner abundances using 4 tiers. Our evaluation differed from that of Clark et al. (2014) in that we evaluated percentile-based benchmarks as WSP benchmarks of biological status and not as sustainable escapement goals based on MSY. In addition, we analyzed performance along gradients in harvest rates and productivity because we found that benchmark performance was most sensitive to these variables, whereas Clark et al. (2014) analyzed tiers according to harvest rates, observed contrast in escapement time-series, and measurement errors in spawning abundances. We found that contrast in escapement was correlated with harvest rates and productivity. When harvest rates were low and productivity was high, contrast was high and percentile benchmarks tended to be higher than "true" value. When harvest rates were high and productivity was low, contrast was low and percentile benchmarks tended to be below "true" values. Where harvest rates and/or productivity are unknown, contrast in escapement time-series could be used as a proxy since low contrast may indicate conditions where percentile benchmarks perform poorly (i.e., fall below "true" benchmarks).

In summary, we evaluated percentile-based benchmarks in retrospective analyses and simulation modelling by comparing their values against "true" underlying benchmark identified by the WSP (Holt et al. 2009), accounting for high uncertainties and possible biases in spawner abundances, and uncertainties in catches, recruitment estimates, and age-at-maturity. We identified conditions where percentile-based benchmarks were the same or higher in value than the "true" benchmarks. In particular, when productivities were moderate to high, the S_{25th} benchmark tended to be higher than the "true" lower benchmark. An alternative lower benchmark, S_{50th} tended to be higher than the "true" lower benchmark at low productivity and low harvest rates, though none of the percentile benchmarks was higher than "true" values at low productivity and moderate to high harvest rates. The upper benchmark, S_{75th}, tended to significantly overestimate the "true" upper benchmark except at high harvest rates and moderate to low productivity. In general, the upper benchmark, S_{50th} better matched the "true" upper benchmark.

4 SUMMARY OF UNCERTAINTIES (OBJECTIVE 2)

Uncertainties in benchmark estimates can arise from uncertainties in underlying spawner abundances and catch data because of sampling methods with low precision (e.g., visual counts) and uncertainties in stock identification. Biases can occur when spawning surveys cover only a portion of the CU, and/or expansion factors are applied which over or underestimate the total abundances. Uncertainties related to sampling errors occur when monitoring of abundances is incomplete over time and/or over locations within CUs. Uncertainties in ages-at-maturity due to incomplete (or lack of) sampling of ages of catch and spawners can result in uncertainties in estimates of recruitment aligned by brood year, which are required for stock-recruitment based benchmarks. Also for stock-recruitment benchmarks, errors-in-variables and time-series biases in parameter estimation can result in biased benchmarks (Walters and Martell 2004, Korman and English 2013). These sources of uncertainty were included in our simulation model evaluating biological benchmarks. Sensitivity analyses covering realistic ranges for parameters revealed two parameters, productivity and harvest rates, had the most significant impacts on benchmark performance. Results from our simulation model were largely robust to the remaining sources of uncertainty.

Status assessments can be generated to reflect underlying uncertainties in benchmarks by considering the probabilities of abundances falling within each status category (Figure 6). We demonstrated an approach for providing probabilistic status assessments using bootstrapped estimates of percentile benchmarks and the posterior distributions of stock-recruitment benchmarks. These probabilistic assessments can help inform a precautionary approach to decision making. For example, a CU may be in the amber zone based on the median benchmark estimate, but given underlying uncertainties the probability of red status may be as high as 49%, warranting precautionary management to keep CUs out of the red zone.

5 ADVICE ON APPLICABILTY OF PERCENTLE-BASED BENCHMARKS (OBJECTIVE 3)

Our third objective was to provide advice on the applicability of percentile-based benchmarks for data-limited CUs of Chum Salmon in southern BC. We assessed applicability relative to stock-recruitment benchmarks identified under the WSP. In general, we found that when productivity was moderate to high, lower and upper benchmarks at the 25th and 50th percentiles of observed spawner abundances tended to be higher than the "true" lower benchmark (S_{gen}) and approximately equivalent to or higher than the "true" upper benchmark (80% S_{MSY}), respectively (Table 6). These percentile benchmarks provide statuses that would tend to be the same or lower than those from "true" underlying benchmarks. We selected these percentile benchmarks, S_{25th} and S_{50th}, for illustrative purposes, but they are aligned with the definitions of benchmarks identified in the WSP. Alternatives could be considered based on the results presented in Figs. 15 and 18, given probabilities of falling below or being far above "true" WSP benchmarks. These results show that an alternative upper benchmark previously recommended for Chum Salmon in BC and applied by Hilborn et al. (2013), S_{75th}, tend to be far higher than the "true" upper benchmark identified under the WSP.

At high harvest rates and moderate productivity, and moderate-high harvest rates and low productivity, none of the lower percentile benchmarks were above the "true" lower benchmark, and in some cases, none of the upper percentile benchmarks were above the "true" upper benchmark. In these cases, percentile-based benchmarks tend to provide statuses that are unrealistically high or optimistic relative to "true" status. Further investigation is required in these cases to identify percentile benchmarks (or other benchmarks) with low risks of over-estimating "true" status. In addition, for the low productivity scenarios, the "true" lower and upper

benchmarks S_{gen} and 80% of S_{MSY}, are close to one another (as documented previously in Holt and Ogden 2013), resulting in the selection of lower and upper percentile benchmarks that match (Table 6, bottom left corner). One consequence of coinciding lower and upper benchmarks is a high probability of dropping below the lower benchmark as CUs are depleted to the upper benchmark, creating conservation and management risks. This risk occurs for both stock-recruitment and percentile benchmarks at low productivity. To manage these risks, information on the probability of status categories (red, amber, green) can be used when making management decisions, since the probability of red status will increase as the CU is depleted near the coinciding lower and upper benchmarks, even if the CU remains in the green zone. Alternatively, the upper benchmark defined by the Wild Salmon Policy and further investigated by Holt et al. (2009) could be reconsidered and set to a higher value (e.g., S_{max}, S_{eq}) that is better distinguished from the lower benchmark, S_{gen}. However, these alternative upper benchmarks, S_{max} and S_{eq}, tend to be much higher than abundances observed in historical records suggesting that green status would rarely be achieved.

Caution should be applied if using S_{25th} and S_{50th} as lower and upper benchmarks without information on productivity and harvest rates. In these cases, declining trends in abundances or low contrast in spawner time-series (<8) would warrant investigation into possible low productivity levels (< 2.5 recruits/spawner) or high harvest pressure (≥40%) where further evaluation of percentile-based benchmarks is warranted. These conditions can result in a shifting baseline whereby percentile benchmarks decline concurrently with abundances, as observed for Southern Coast Streams CU (Fig.5a,iii). Alternatively, if productivity can be estimated through hierarchical modelling or meta-analyses, then benchmarks can be applied accordingly. Where exploitation rates are estimated to be high on average over the time-series and productivity is unknown, then the risks of percentile-based benchmarks providing over-estimates of status (i.e., being overly optimistic) are relatively high, and alternative percentile benchmarks should be evaluated.

Further caution is required if CUs are severely depleted far below "true" benchmarks, possibly resulting in depensation due to increased predation or reduced mating success at very low abundances. While our simulations were initialized at 20% of capacity, the subsequent time-series tended to cover a wide range of spawner abundances. Depensatory effects at low spawner abundances (not modelled here) may reduce productivity preventing abundances from rebounding to high levels. In addition, where spawner time-series are short and highly variable over time, percentile benchmarks may also be highly variable (e.g., Southern Coastal Streams, Figure 5a,iii, 1960-1980), though these benchmarks tend to stabilize as time-series exceed 30 years.

Here we assumed stock-recruitment benchmarks adequately captured "true" underlying benchmarks of status, defined by the WSP. The Ricker model underlying the stock-recruitment benchmarks assumes density dependence in recruitment at moderate to high spawner abundances within CUs. However, for Chum Salmon in BC, the spatial scale of density dependence is uncertain, and may differ from the scale of CUs (e.g., occurring at local spawning sites within CUs, or at a larger multi-CU level since Chum Salmon fry from numerous CUs can overlap in estuaries). Despite uncertainty in the underlying scale of density dependence, the Ricker model was found to adequately capture the empirical relationship between spawner and recruits for Chum Salmon in BC, Alaska, and Washington State (Dorner et al. 2008), and was used here.

Alternative approaches for identifying benchmarks using habitat or watershed characteristics to estimate the capacity of the system to sustain spawners and juveniles could help validate percentile benchmarks given the caveats in percentile approaches described above. Where percentile benchmarks are deemed unreliable, habitat-based benchmarks of capacity could be

used as an alternative as applied to Chinook Salmon in BC (Parken et al. 2006), and to Fraser River Sockeye Salmon to inform prior probability distributions of S_{max} for stock-recruitment benchmarks (Grant et al. 2011).

Integrated assessments under Strategy 1 of the WSP consider multiple metrics and benchmarks of status on abundances, trends in abundance, distribution, and fishing mortality relative to productivity, in part to account for the shortcomings and uncertainties in each individual metrics or benchmarks such as those identified here for percentile benchmarks. For example, metrics on the distribution of spawning across sub-populations can supplement those on abundances by indicating when widely distributed CUs have become more concentrated over time, possibly resulting in a loss of resilience not reflected in metrics on abundances (Holt et al. 2009). In addition, benchmarks on long-term trends in abundances over time have been identified at 25% and 50% of the long-term geometric median of observed spawner abundances. Those benchmarks are lower than the percentile benchmarks identified here, and were identified based on expert opinion not quantitative analyses (Holt et al. 2009). Our percentile benchmarks provide status assessments that are more consistent with stock-recruitment based benchmarks on abundances, and could be considered in place of current benchmarks on long-term trends in integrated assessments.

Furthermore, when applied to large-scale status assessments of CUs region-wide, percentile benchmarks may be valuable for providing preliminary snap-shots of status for prioritizing work on monitoring and assessment, in the absence of integrated status assessments. CUs or groups of CUs that are observed to be below lower percentile benchmarks may trigger further work into status assessments and/or monitoring.

Our analyses evaluating sensitivity of benchmark performance to various levels of productivity were in the context of average, long-term values, and not interannual variability in productivity. We recommend that benchmarks be chosen to reflect long-term trends, instead of adjusting them annually in response to yearly changes in productivity.

Finally, our analyses focus on Chum Salmon, though similar analyses could be applied to other species. Differences in data quality and quantity (e.g., of spawner time-series, catches or exploitation rates, and age-at-maturity), and life-history characteristics (e.g., variability in age-at-maturity, recruitment autocorrelation, and productivity) may affect relative performance of benchmarks limiting our ability to provide recommendations on other species presently.

6 ACKNOWLEDGEMENTS

7 REFERENCES

Butterworth, D. S. 2007. Why a management procedure approach? Some positives and negatives. ICES J. Mar. Sci. 64:613-617.

Clark, R. A., D. M. Eggers, A. R. Munro, S. J. Fleischman, B. G. Bue, and J. J. Hasbrouck. 2014. An evaluation of the percentile approach for establishing sustainable escapement goals in lieu of stock productivity information. Alaska Dept. Fish and Game Fish. Man. No. 14-06.

Collie, J. S., R. M. Peterman, B. M. Zuehlke, and Y. Chen. 2012. A fisheries risk-assessment framework to evaluate trade-offs among management options in the presence of time-varying productivity. Can. J. Fish. Aquat. Sci. 69:209-223.

Cousens, N. B. F., G. A. Thomas, C. G. Swann, and M. C. Healey. 1982. A review of salmon escapement estimation techniques. Can. Tech. Rep. Fish. Aquat. Sci. No. 1108:122pp.

Debertin, A. J., J. R. Irvine, C. A. Holt, G. Oka, and M. Trudel. 2016. Marine growth patterns of southern British Columbia chum salmon explained by interactions between density-dependent competition and changing climate. Can. J. Fish. Aquat. Sci. 74:1077–1087.

DFO. 2005. Canada's Policy for the Conservation of Wild Pacific Salmon. Vancouver, British Columbia.

DFO. 2015. Wild salmon policy biological status assessment for conservation units of interior Fraser River Coho Salmon (*Oncorhynchus kisutch*). DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2015/005.

DFO. 2016. Integrated Biological Status of Southern British Columbia Chinook Salmon (*Oncorhynchus tshawytscha*) Under the Wild Salmon Policy. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2016/042.

Dobson, D., D. O'Brien, and G. Pestal. 2009. Certification Unit Profile: West Coast Vancouver Island Chum Salmon. Can. Man. Rep. Fish. Aquat. Sci. No. 2881.

Dorner, B., K. R. Holt, R. M. Peterman, C. Jordan, D. P. Larsen, A. R. Olsen, and O. I. Abdul-Aziz. 2013. Evaluating alternative methods for monitoring and estimating responses of salmon productivity in the North Pacific to future climatic change and other processes: A simulation study. Fish. Res. 147:10-23.

Dorner, B., R. M. Peterman, S. L. Haeseker, and B. J. Pyper. 2008. Reconstructed historical trends in productivity of 120 Pacific pink, chum, and sockeye salmon populations by using a Kalman filter. Can. J. Fish. Aquat. Sci. 65:1842–1866.

English, K., G. Ruggerone, and S. Devitt. 2014. Surveillance report - British Columbia Chum Salmon fisheries. Intertek Fisheries Certification Ltd.

Fleischman, S. J., M. J. Catalano, R. A. Clark, D. R. Bernard, and Y. Chen. 2013. An age-structured state-space stock–recruit model for Pacific salmon (*Oncorhynchus* spp.). Can. J. Fish. Aquat. Sci. 70:401-414.

Godbout, L., J.R. Irvine, D. Bailey, P. Van Will, and C. McConnell. 2004. Stock status of wild Chum Salmon (*Oncorhynchus keta Walbaum*) returning to British Columbia's central and Johnstone and Georgia Straits (excluding the Fraser River). DFO Can. Sci. Advis. Sec. Res. Doc. 2004/007. vi + 45 p.

Godbout, L., C. Fu, and J. R. Irvine. 2006. Evaluation of Chum Salmon fishery performance using Ricker and Beverton–Holt stock recruitment approaches in a Bayesian framework. American Fisheries Society Symposium 49:1417-1433.

Grant, S., and G. Pestal. 2013. Integrated Biological Status Assessments Under the Wild Salmon Policy Using Standardized Metrics and Expert Judgement: Fraser River Sockeye Salmon (*Oncorhynchus nerka*) Case Studies. DFO Can. Sci. Advis. Sec. Res. Doc. 2012/106. v + 132 p.

Grant, S. C. H., B. L. MacDonald, T. E. Cone, C. A. Holt, A. Cass, E. J. Porszt, J. M. B. Hume, and L. B. Pon. 2011. Evaluation of uncertainty in Fraser Sockeye WSP Status using abundance and trends in abundance metrics. DFO Can. Sci. Advis. Sec. Res. Doc. 2011/087. viii + 183 p.

Groot, C., and L. Margolis. 1991. Pacific Salmon Life Histories. University of British Columbia Press, Vancouver, B.C.

Hilborn, R., D. Schmidt, K. English, and S. Devitt. 2013. British Columbia Chum Salmon (*Oncorhynchus keta*) Fisheries: British Columbia Coastal and Adjacent Canadian Pacific EEZ Waters, Public Certification Report. Submitted to Canadian Pacific Sustainable Fisheries Society. Report submitted to Canadian Pacific Sustainable Fisheries Society.

Hilborn, R., and C. J. E. Walters. 1992. Quantitative Fisheries Stock Assessment; Choice Dynamics & Uncertainty. Springer US.

Holt, C. A. 2009. Evaluation of Benchmarks for Conservation Units in Canada's Wild Salmon Policy: Technical Documentation. DFO Can. Sci. Advis. Sec. Res. Doc. 2009/059. x + 50 p.

Holt, C. A., and M. J. Bradford. 2011. Evaluating benchmarks of population status for Pacific salmon. N. Amer. J. Fish. Manag.31:363-378.

Holt, C. A., A. Cass, B. Holtby, and B. Riddell. 2009. Indicators of Status and Benchmarks for Conservation Units in Canada's Wild Salmon Policy. DFO Can. Sci. Advis. Sec. Res. Doc. 2009/058. viii + 74 p.

Holt, C. A., and M. J. P. Folkes. 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. Fish. Res. 171:188-200.

Holt, C. A., and J. R. Irvine. 2013. Distinguishing benchmarks of biological status from management reference points: A case study on Pacific salmon in Canada. Envir. Cons. 40:345-355.

Holt, C. A., and A. Odgen. 2013. Software for assessing status of Conservation Units under Canada's Wild Salmon Policy: Instructional manual. Can. Tech. Rep. Fish. Aquat. Sci. 3058:vi + 43 p.

Holt, C. A., and R. M. Peterman. 2006. Missing the target: implementation uncertainties in fisheries on Fraser River, British Columbia, sockeye salmon (*Oncorhynchus nerka*). Can. J. Fish. Aquat. Sci. 63:2722-2733.

Holtby, L. B., and K. A. Ciruna. 2007. Conservation units for Pacific salmon under the Wild Salmon Policy. DFO Can. Sci. Advis. Sec. Res. Doc. 2007/070. viii + 350 p.

Korman, J., and K. English. 2013. Benchmark Analysis for Pacific Salmon Conservation Units in the Skeena Watershed. Report prepared for the Pacific Salmon Foundation.

Korman, J., R. M. Peterman, and C. J. Walters. 1995. Empirical and theoretical analyses of correction of time series bias in stock-recruitment relationships of sockeye salmon (*Oncorhynchus nerka*). Can. J. Fish. Aquat. Sci. 52:2174-2189.

Malick, M. J., and S. P. Cox. 2016. Regional-Scale Declines in Productivity of Pink and Chum Salmon Stocks in Western North America. PLoS One 11:e0146009.

Malick, M. J., S. P. Cox, F. J. Mueter, B. Dorner, and R. M. Peterman. 2017. Effects of the North Pacific Current on the productivity of 163 Pacific salmon stocks. Fisheries Oceanography 26:268-281.

McElhany, P., M. Rucklelshaus, M. Ford, T. Wainwright, and E. Bjorkstedt. 2000. Viable Salmonid Populations and the Recovery of Evolutionary Significant Units., U.S. Dept. Commer.

Morris, D. J., D. C. Speirs, A. I. Cameron, and M. R. Heath. 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.

Parken, C. K., R. E. McNicol, and J. R. Irvine. 2006. Habitat-based methods to estimate escapement goals for data limited Chinook salmon stocks in British Columbia, 2004. DFO Can. Sci. Advis. Sec. Res. Doc. 2006/083. vii + 67 p.

Peterman, R. M. 2004. Possible solutions to some challenges facing fisheries scientists and managers. ICES J. Mar. Sci. 61:1331-1343.

Peterman, R. M., B. J. Pyper, and B. MacGregor. 2003. Use of the Kalman filter to reconstruct historical trends in productivity of Bristol Bay sockeye salmon (*Oncorhynchus nerka*). Can. J. Fish. Aquat. Sci. 60:809-824.

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

Portley, N., and H. J. Geiger. 2014. Limit Reference Points for Pacific Salmon Fisheries. N. Amer. J. Fish. Manag. 34:401-410.

Pujol, G., B. Iooss, A. Janon, w. c. f. S. D. Veiga, J. Fruth, L. Gilquin, J. Guillaume, L. L. Gratiet, P. Lemaitre, B. Ramos, and T. Touati. 2015. sensitivity: Sensitivity Analysis. R package version 1.11.1.

Pyper, B. J., F. J. Mueter, and R. M. Peterman. 2002. Spatial covariation in survival rates of northeast Pacific chum salmon. Trans. Amer. Fish. Soc. 131:343-363.

R Development Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ricker, W.E., 1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Board Can. 191, 382 p.

Schnute, J. T., and A. R. Kronlund. 1996. A management oriented approach to stock recruitment analysis. Can. J. Fish. Aquat. Sci. 53:1281-1293.

Su, Y. S., and M. Yajima. 2012. R2jags: A Package for Running jags from R. R package version 0.03-08.

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

Szerlong, R. G., and D. E. Rundio. 2008. A statistical modeling method for estimating mortality and abundance of spawning salmon from a time series of counts. Can. J. Fish. Aquat. Sci. 65:17-26.

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

Walters, C. J., and S. J. D. Martell. 2004. Fisheries Ecology and Management. Princeton University Press, Princeton.

8 TABLES

Table 1. Current benchmark values across three methods used: standard Ricker model, hierarchical Ricker model (S_{gen} and 80% S_{MSY}) and percentiles (25th and 75th).

Southern Coastal Streams

Method	Standard Ricker	Hierarchical Ricker	Percentile
Lower Benchmark	9,636	9,994	5,425
Upper Benchmark	10,372	11,711	54,350

Northeast Vancouver Island

Method	Standard Ricker	Hierarchical Ricker	Percentile
Lower Benchmark	16,506	16,292	16,519
Upper Benchmark	18,503	19,494	75,136

Upper Knight

Method	Standard Ricker	Hierarchical Ricker	Percentile
Lower Benchmark	2,991	3,086	2,006
Upper Benchmark	4,600	4,572	11,191

Loughborough

Method	Standard Ricker	Hierarchical Ricker	Percentile
Lower Benchmark	12,002	12,316	17,313
Upper Benchmark	18,219	18,301	46,303

Bute Inlet

Method	Standard Ricker	Hierarchical Ricker	Percentile
Lower Benchmark	20,528	21,155	11,275
Upper Benchmark	33,752	33,247	85,517

Georgia Strait

Method	Standard Ricker	Hierarchical Ricker	Percentile
Lower Benchmark	91,724	113,305	202,269
Upper Benchmark	187,546	201,020	445,139

Howe Sound to Burrard Inlet

Method	Standard Ricker	Hierarchical Ricker	Percentile
Lower Benchmark	97,554	107,571	85,394
Upper Benchmark	171,126	177,421	303,280

Northwest Vancouver Island

Method	Standard Ricker	Hierarchical Ricker	Percentile
Lower Benchmark	11,995	11,997	24,811
Upper Benchmark	19,871	19,921	73,650

Southwest Vancouver Island

Method	Standard Ricker	Hierarchical Ricker	Percentile
Lower Benchmark	65,109	66,202	204,065
Upper Benchmark	120,726	121,273	433,640

Table 2. Parameter and benchmark estimates, lower credible limits (LCL)and upper credible limits (UCL) delineated as 2.5th and 97.5th posterior densities for the most recent year.

South Coast Streams

Model	Standard			Hierarchical		
Statistic	Estimate	LCL	UCL	Estimate	LCL	UCL
Ricker α	1.39	0.92	2.23	1.60	1.02	2.43
S_{max}	80,275	43,303	218,015	67,219	41,050	183,913
S_{gen}	9,636	2,131	19,263	9,994	4,292	18,702
80% S_{MSY}	10,372	1,734	20,370	11,711	3,647	20,687

Northeast Vancouver Island

Model	Standard			Hierarchical		
Statistic	Estimate	LCL	UCL	Estimate	LCL	UCL
Ricker α	1.53	1.05	2.23	1.70	1.14	2.43
S_{max}	115,696	68,614	313,299	101,040	63,165	284,140
S_{gen}	16,506	7,927	31,386	16,292	10,155	31,635
80% S_{MSY}	18,503	6,688	33,262	19,494	9,098	34,620

Upper Knight

Model	Standard			Hierarchical		
Statistic	Estimate	LCL	UCL	Estimate	LCL	UCL
Ricker α	2.22	1.19	4.04	2.18	1.34	3.56
S_{max}	16,523	9,410	62,829	16,756	9,804	57,175
S_{gen}	2,944	1,485	9,167	3,089	1,736	9,365
80% S_{MSY}	4,581	1,980	11,619	4,578	2,429	12,270

Loughborough

Model	Standard			Hierarchical		
Statistic	Estimate	LCL	UCL	Estimate	LCL	UCL
Ricker α	2.30	1.59	3.23	2.24	1.66	3.06
S_{max}	62,730	43,696	123,151	64,033	44,811	116,832
S_{gen}	11,992	8,095	21,103	12,227	8,440	20,857
80% S_{MSY}	18,401	13,671	27,404	18,194	13,628	27,109

Bute Inlet

Model	Standard			Hierarchical		
Statistic	Estimate	LCL	UCL	Estimate	LCL	UCL
Ricker α	2.46	1.62	3.77	2.32	1.64	3.44
S_{max}	106,264	69,179	246,135	111,430	73,740	278,742
S_{gen}	20,222	12,203	44,095	21,257	13,489	47,248
80% S_{MSY}	33,348	23,044	60,030	33,484	22,782	62,854

Georgia Strait

Model	Standard			Hierarchical		
Statistic	Estimate	LCL	UCL	Estimate	LCL	UCL
Ricker α	3.08	2.05	4.77	2.67	1.97	4.19
S_{max}	493,198	301,072	1,083,934	608,911	336,236	1,143,835
S_{gen}	90,983	44,872	206,206	116,883	53,905	216,737
80% S_{MSY}	186,802	141,661	303,619	203,327	146,895	313,297

Howe Sound to Burrard Inlet

Model	Standard			Hierarchical		
Statistic	Estimate	LCL	UCL	Estimate	LCL	UCL
Ricker α	2.63	1.81	3.75	2.47	1.79	3.49
S_{max}	511,173	308,310	1,657,672	559,155	333,798	1,837,905
S_{gen}	97,554	54,892	310,845	107,571	60,229	344,097
80% S_{MSY}	171,126	119,131	410,187	177,421	120,094	453,338

North West Coast Vancouver Island

Model	Standard			Hierarchical		
Statistic	Estimate	LCL	UCL	Estimate	LCL	UCL
Ricker α	2.50	1.66	3.75	2.51	1.65	3.76
S_{max}	62,398	44,590	108,771	62,597	44,742	110,640
S_{gen}	11,995	7,754	18,845	11,997	7,724	18,841
80% S_{MSY}	19,871	15,779	25,247	19,921	15,671	25,492

South West Coast Vancouver Island

Model	Standard			Hierarchical		
Statistic	Estimate	LCL	UCL	Estimate	LCL	UCL
Ricker α	2.81	1.69	4.55	2.78	1.64	4.56
S_{max}	344,426	232,644	731,855	348,571	235,912	746,969
S_{gen}	65,109	35,981	124,899	66,202	36,558	125,382
80% S_{MSY}	120,726	99,282	161,752	121,273	99,754	160,359

Table 3. 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. Probability of each status is also given, which is estimated using the posterior densities for the Ricker models, and the bootstrapped status for percentile benchmarks.

Conservation Unit	Percentile Status	Standard Ricker Status	Hierarchical Ricker Status
South Coast Streams	Red (61%)	Red (93%)	Red (98%)
Northeast Vancouver Island	Amber (100%)	Green (90%)	Green (90%)
Upper Knight	Amber (96%)	Green (82%)	Green (81%)
Loughborough	Amber (89%)	Green (67%)	Green (69%)
Bute Inlet	Amber (95%)	Green (95%)	Green (96%)
Georgia Strait	Green (90%)	Green (100%)	Green (100%)
Howe Sound to Burrard Inlet	Green (100%)	Green (100%)	Green (100%)
North West Coast Vancouver island	Green (69%)	Green (100%)	Green (100%)
South West Coast Vancouver Island	Red (96%)	Green (99%)	Green (99%)

Table 4. Proportion of years where percentile-based status and Ricker-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 were more precautionary than Ricker-based status (i.e., had lower status).

Conservation unit	Percentile-based status match standard Ricker-based status	Percentile- based status match hierarchical Ricker-based status	Percentile-based status match or more precautionary than standard Ricker-based status	Percentile-based status match or more precautionary than hierarchical Ricker-based status
South Coast Streams	0.35	0.47	1.00	1.00
Northeast Vancouver Island	0.42	0.34	1.00	1.00
Upper Knight	0.50	0.50	0.78	0.78
Loughborough	0.23	0.23	1.00	1.00
Bute Inlet	0.60	0.58	0.75	0.73
Georgia Strait	0.35	0.35	1.00	1.00
Howe Sound to Burrard Inlet	0.58	0.60	1.00	1.00
North West Coast Vancouver island	0.14	0.15	1.00	1.00
South West Coast Vancouver Island	0.00	0.00	1.00	1.00

Table 5. Parameters used in simulation model for the base case and sensitivity analyses.

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 (recruits/spawner at low spawner abundances)	2.72 (=1 $\log_e(\alpha)$)	1.64 (low) and 7.39 (high)	1.64-7.39 (=0.5-2.0 $\log_e(\alpha)$)
	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, ϖ, 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, ϖ_{pr}, specified in a multivariate logistic distribution	0.1	0 (low) and 0.9 (high)	0.1-0.9
	Standard deviation in observation errors of spawners	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 (neg. 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	20%	10% (low) and 60% (high)	10%-60%
	Outcome uncertainty (standard deviation in differences between target and realized harvest rates)	0.3	0.5 (high)	0-0.9

*Table 6. Selected percentile-based lower and upper benchmarks identified to be similar or higher in value than stock-recruitment based benchmarks under the WSP, along gradients in productivity (recruits/spawner) and average harvest rates. * denotes the low-productivity scenario where lower and upper Ricker-based benchmarks are very close to one another, resulting in lower and upper percentile-based benchmarks that are the same.*

		Harvest rate		
		<20%	≥20 and <40%	≥40% and <60%
Productivity	>4	25th (lower) 50th (upper)	25th (lower) 50th (upper)	25th (lower) 50th (upper)
	>2.5 and ≤4	25th (lower) 50th (upper)	25th (lower) 50th (upper)	Further evaluation required
	≤2.5 and >1.5	*50th (lower and upper)	Further evaluation required	Further evaluation required

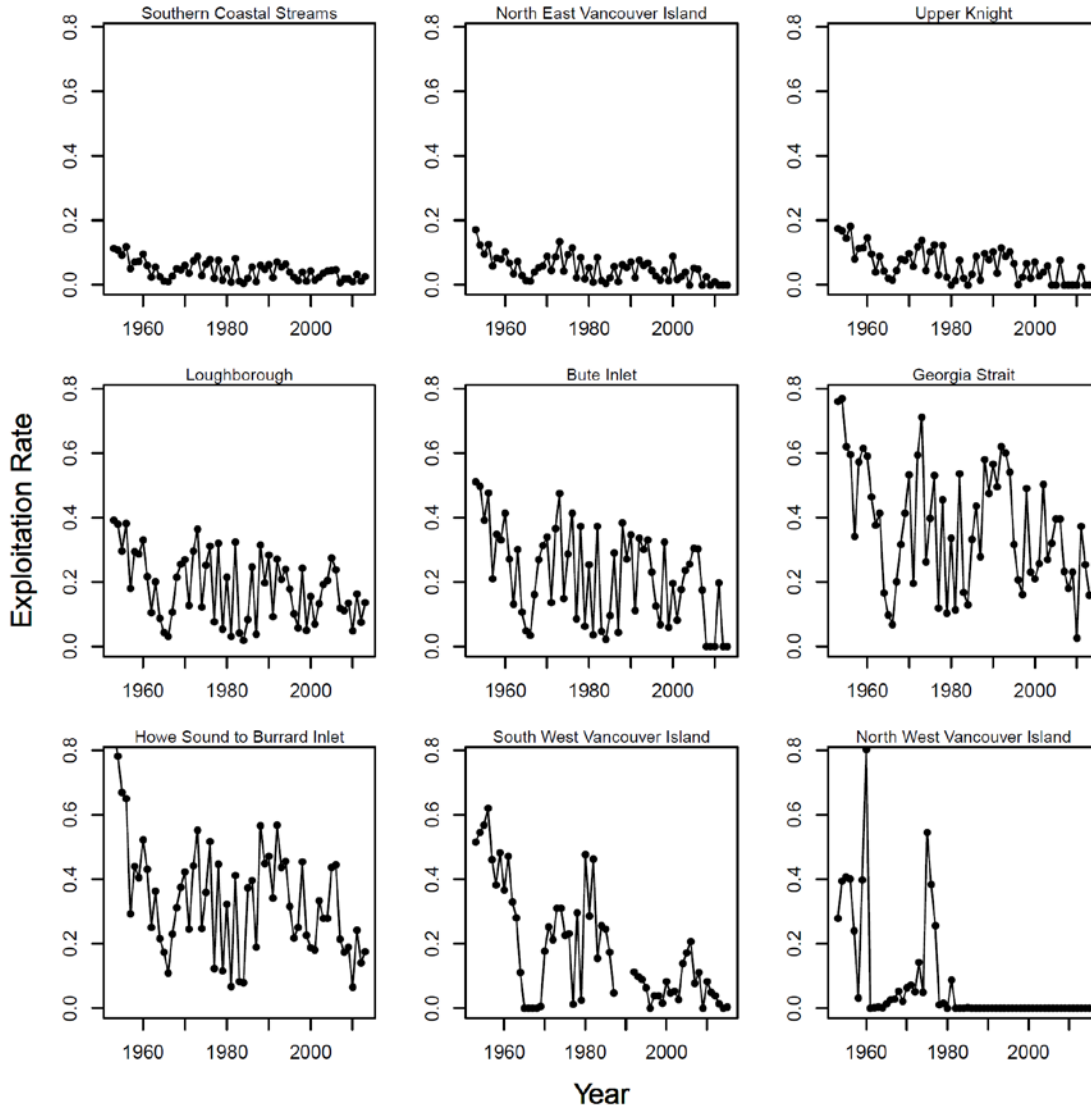# APPENDIX A. EXPLOITATION RATES

Figure A1. Estimated exploitation rates over time, across 9 Chum Salmon CUs on Inner South Coast of BC and West Coast of Vancouver Island.

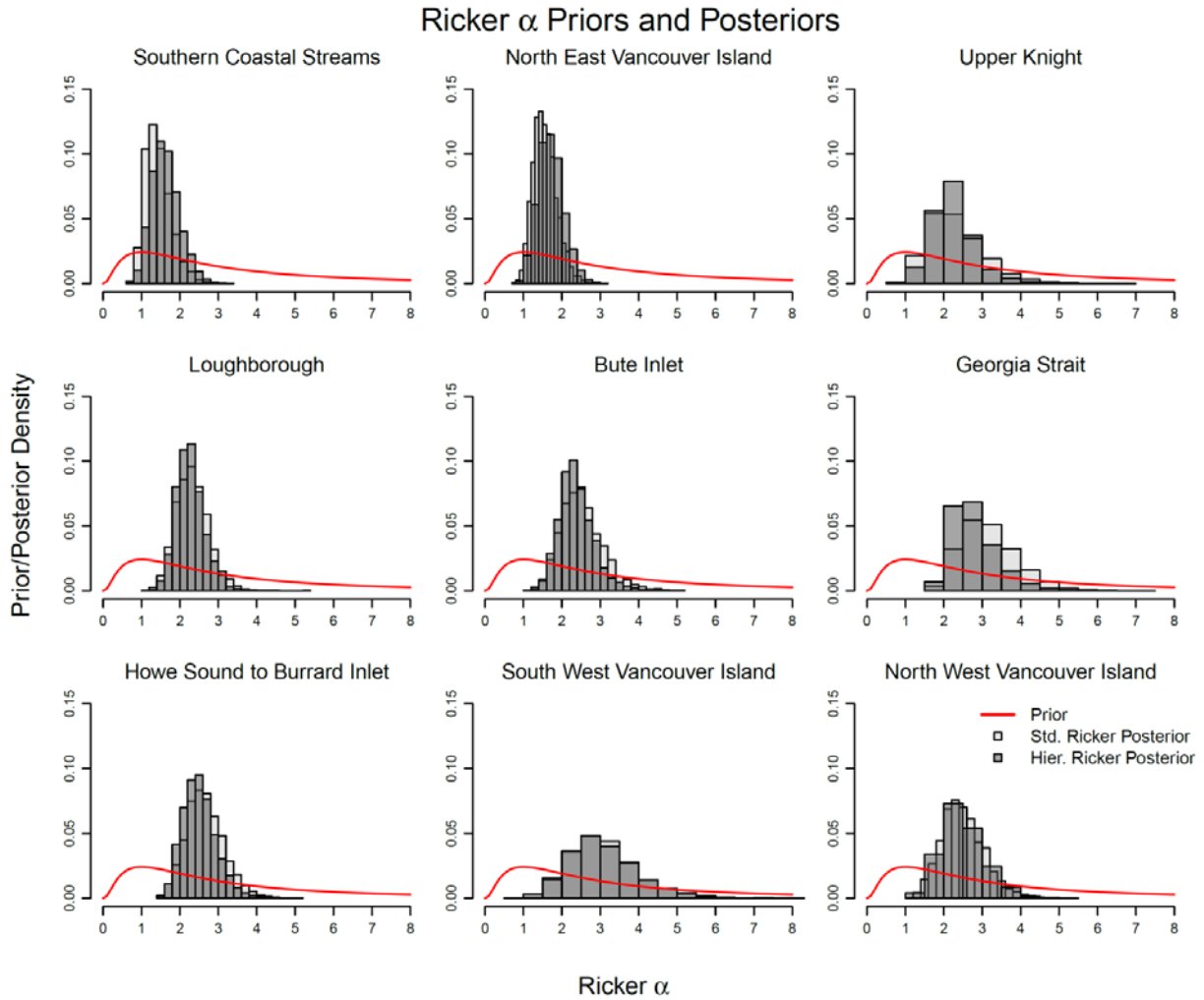# APPENDIX B. PRODUCTIVITY PRIORS AND POSTERIORS

Figure B1. Priors and posteriors for Ricker α parameters for both the standard and hierarchical Ricker models.

APPENDIX C. SENSITIVITY ANALYSIS ON PRIORS

Two prior formulations on Ricker β, via its reciprocal: S_{max}, were used, as described in equations C1 and C2:

(C1) $S_{max} \sim uniform(1, max(S_{obs}) \cdot 2)$

(C2) $S_{max} \sim lognormal(log(mean(S_{obs})), 1/\tau_S), \quad \tau_S = 1/\log(CV^2 + 1)$

For the parameterization of the uniform prior we assumed that S_{max} was less than twice the maximum observed spawner value (S_{obs}), 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 Smax as the uniform distribution. The lognormal prior is weakly informative, as it pulls posterior distributions of Smax 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 Smax 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.

For all CUs, 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 (Fig. C1). Furthermore, when comparing statuses, models fit with either prior matched between 88-100% of years depending on CU, and therefore do not appear to make a significant difference in the assignment of status
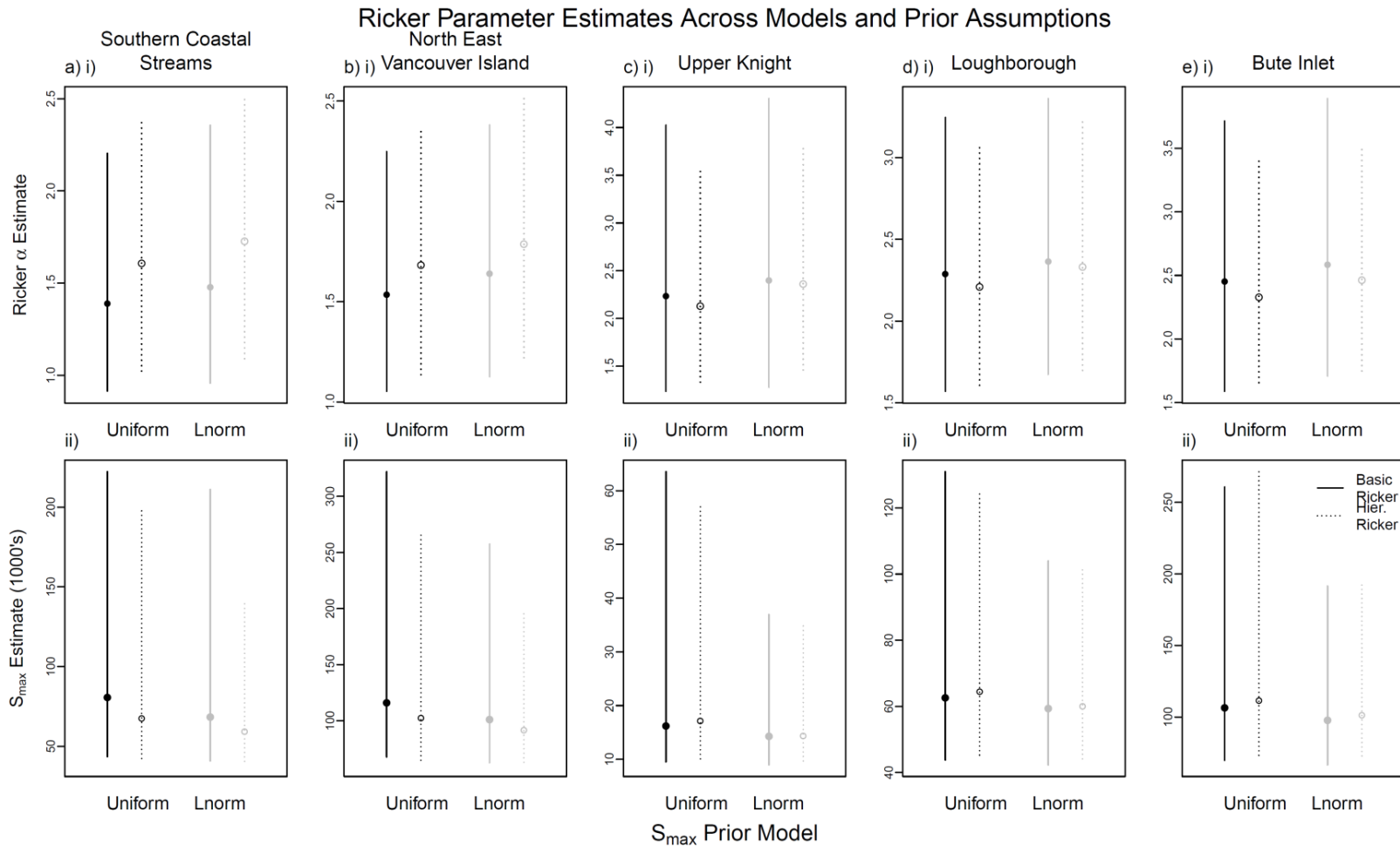
Figure. C1. Model estimates for Ricker α (top row) and S_{max} (bottom row) across prior distributions (uniform and lognormal; see appendix B) for S_{max} and standard and hierarchical Ricker model structures (solid and dotted lines, respectively) for each CU in the final year (2012 for ISC CUs, 2016 for WCVI CUs). Circles indicate posterior medians, and lines indicate 95% credible intervals of estimates.
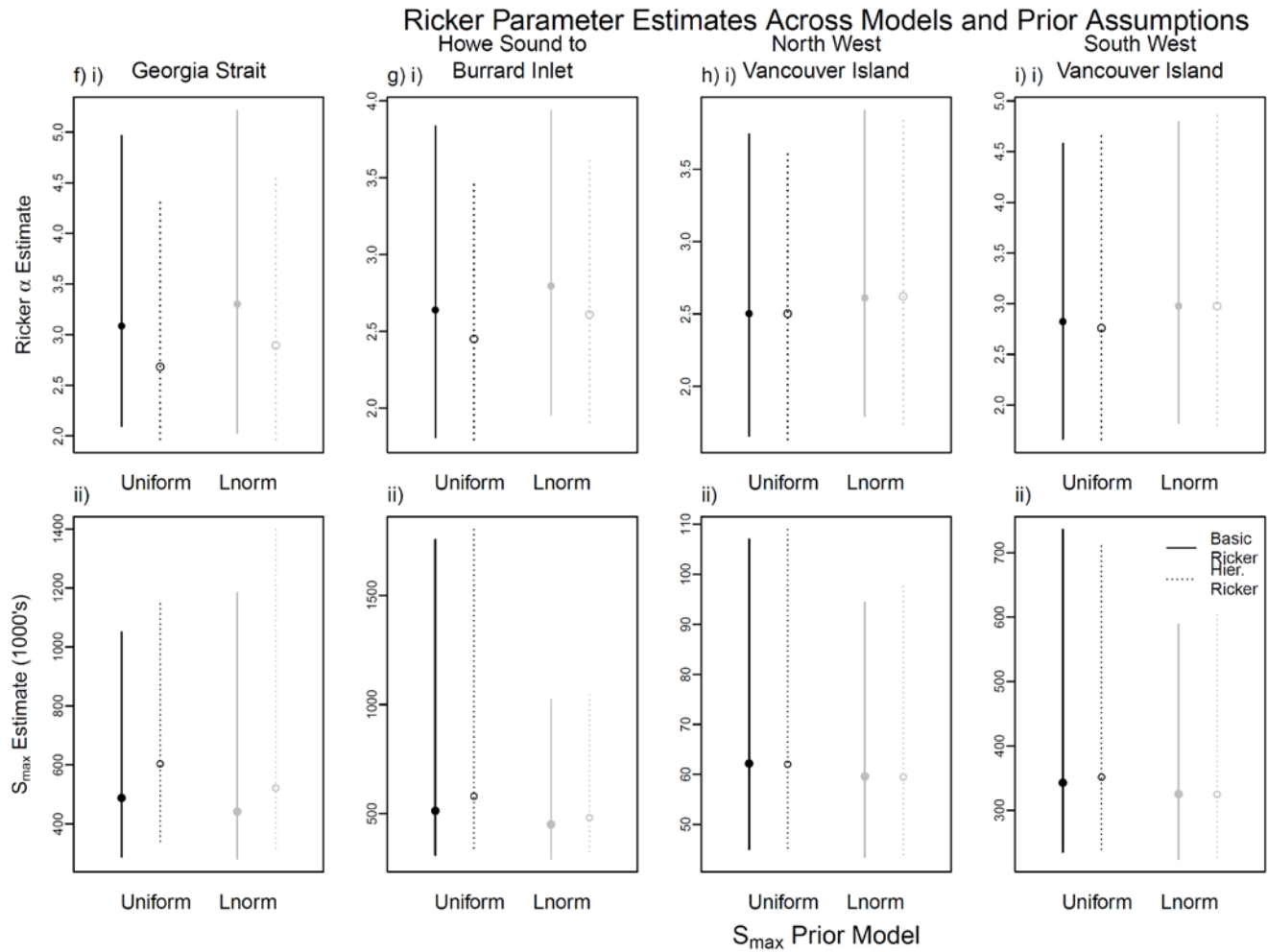
Figure C1 continued.

APPENDIX D. FINAL YEAR RICKER CURVES
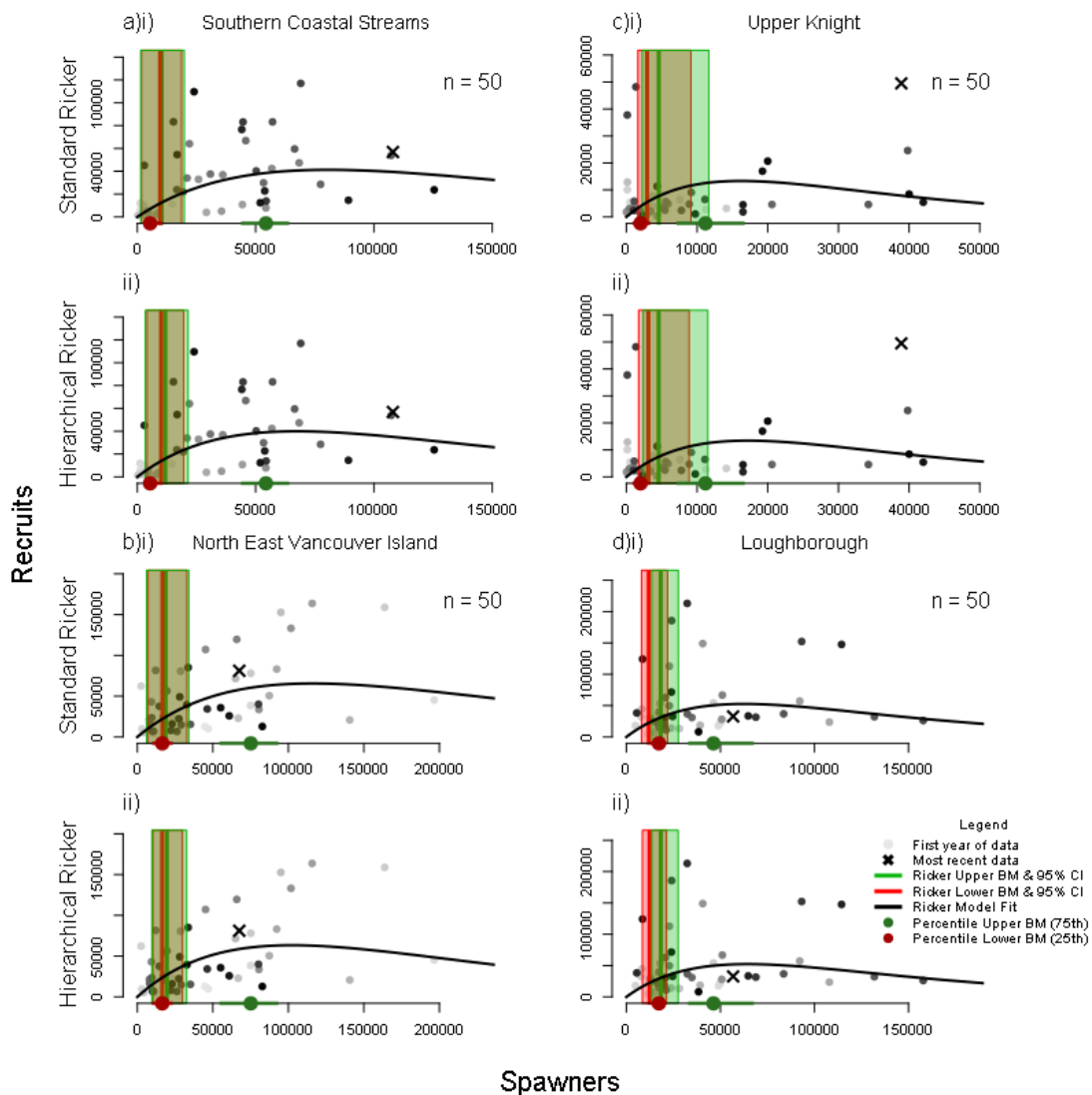
Figure. D1 Observed spawner-recruit data for 9 CUs, Southern Coastal Streams (a), North East Vancouver Island (b), Upper Knight (c), Loughborough (d), Bute Inlet (e), Georgia Strait (f), Howe Sound to Burrard Inlet (g), South West Vancouver Island (h), and North West Vancouver Island (i), with fitted Ricker curves and associated benchmarks for (i) the standard Ricker model, and (ii) the hierarchical Ricker model. Shaded regions indicate 95% credible intervals. 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 for ISC CUs (a-g), 2010 for WCVI CUs (h,i). Colours of points increase in darkness as years progress towards the current year.

Final Year Ricker Curves and Benchmarks
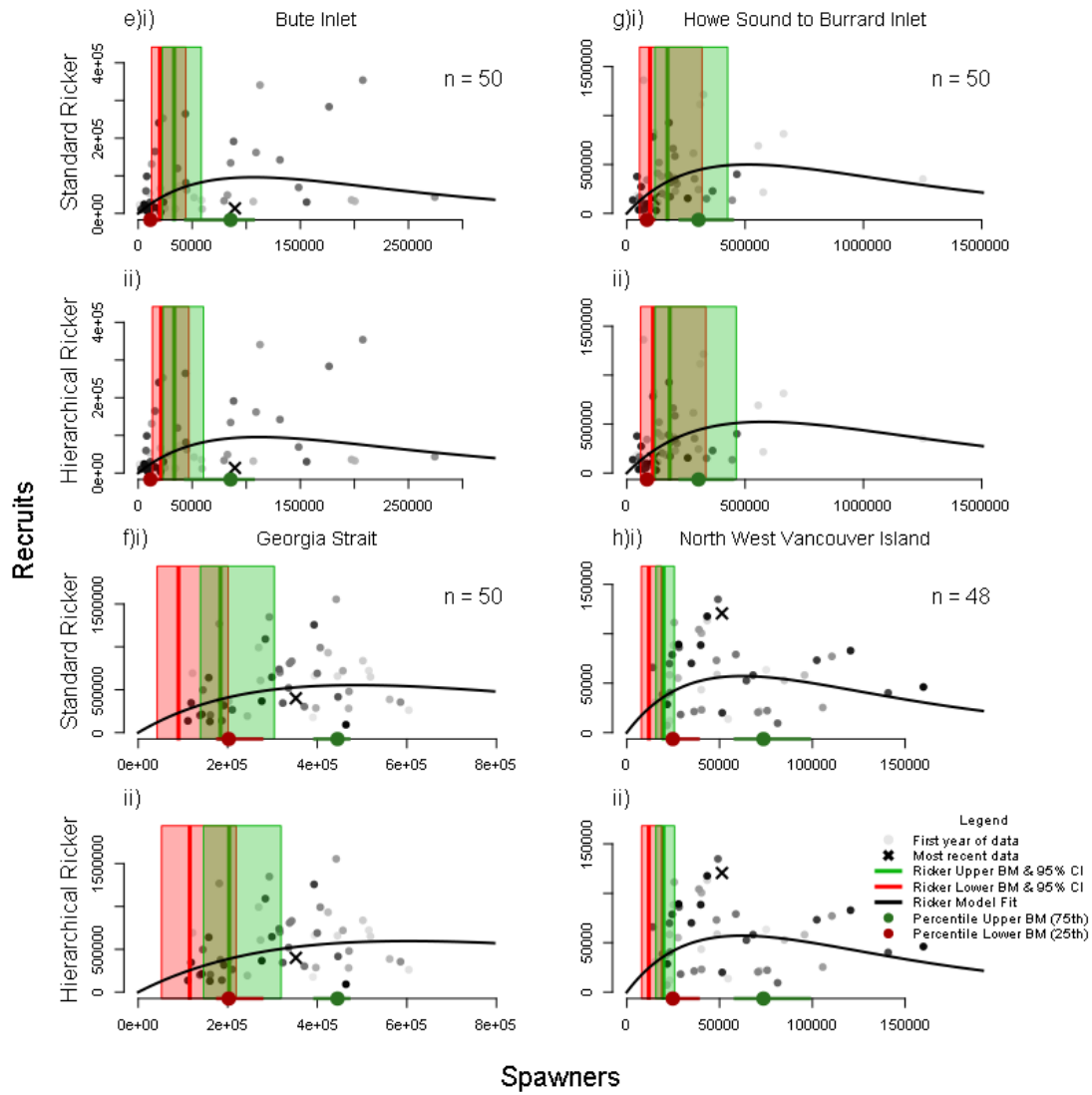
Figure D1 continued.

Final Year Ricker Curves and Benchmarks
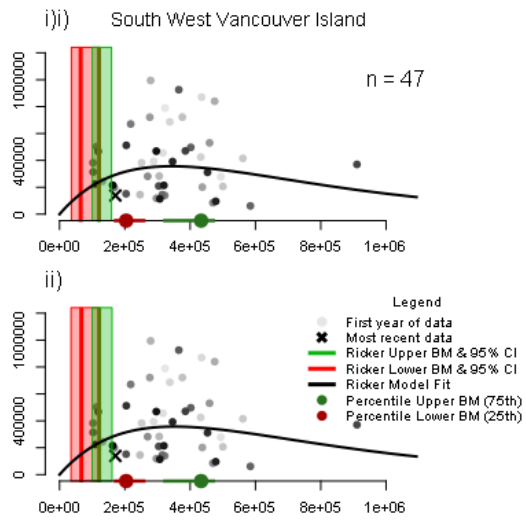
Figure D1 continued.

Table D1. Model fit (r^2 between observed and predicted recruitment) for standard and hierarchical Ricker models.

CU	r^2	
	Standard Ricker	Hierarchical Ricker
South Coast Streams	0.47	0.47
Northeast Vancouver Island	0.29	0.27
Upper Knight	0.00	0.00
Loughborough	0.07	0.07
Bute Inlet	0.06	0.07
Georgia Strait	0.00	0.00
Howe Sound to Burrard Inlet	0.00	0.01
North West Coast Vancouver island	0.03	0.02
South West Coast Vancouver Island	0.00	0.00

APPENDIX E. RETROSPECTIVE PERFORMANCE AGAINST PRODUCTIVITY

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. E1) 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 upper benchmark ($80\% \, S_{MSY}$), while at low/moderate productivities, percentile lower benchmarks can dip below S_{gen}. There is a curvilinear relationship between and the Ricker α parameter and these ratios. For the lower benchmarks, higher productivity is associated with percentile benchmarks that are relatively high compared to Ricker based benchmarks. While for upper benchmarks, percentile and Ricker-based upper benchmarks tend to be most similar at intermediate productivity levels.
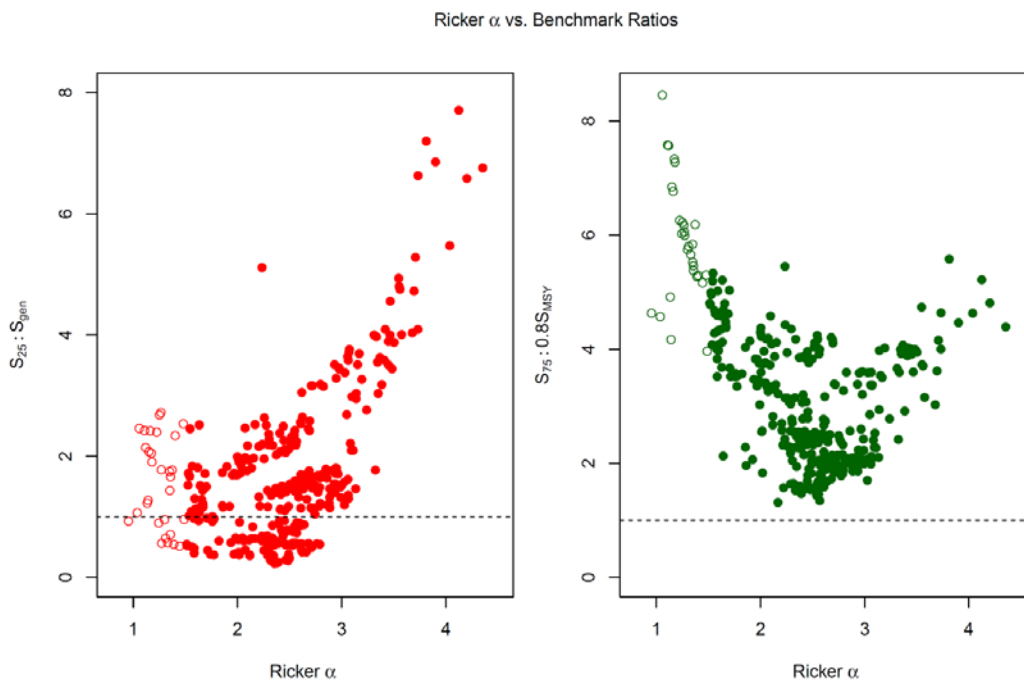
Figure. E1. Ricker α (productivity) parameters vs. ratios of percentile-based benchmarks to Ricker-based benchmarks for ISC and WCVI CUs. Left plot shows ratio for lower benchmarks ($S_{25} : S_{gen}$), right plot shows ratio for upper benchmarks ($S_{75} : 0.8 S_{MSY}$). Points lying above the dashed line identify cases where the percentile benchmark is larger than the Ricker-based benchmark. The empty circles indicate points with $\alpha < 1.5$, which would not have been used to assess status (Holt and Ogden 2013).

APPENDIX E REFERENCES

Holt, C.A., and Ogden, A. 2013. Software for assessing status of Conservation Units under Canada's Wild Salmon Policy: Instructional manual. Can. Tech. Rep. Fish. Aquat. Sci. 3058: vi + 43 p.

APPENDIX F. MODEL EQUATIONS

POPULATION DYNAMICS SUB-MODEL

Our model simulated the dynamics of a single, hypothetical CU of Chum Salmon with 5 sub-populations. Chum Salmon spawn as adults and rear as juveniles in freshwater. Juveniles then migrate as smolts from freshwater to the ocean where they mature before returning to their natal spawning grounds. In Southern BC, Chum Salmon usually return to freshwater as adults at age 3, 4, or 5, and spawn only once before dying. The number of adult recruits that return to freshwater in year, t, in subpopulation j, $R_{t,j}$, was calculated from the total number of adult recruits generated from spawners 3, 4, and 5 years prior, $R'_{t-3,j}$, $R'_{t-4,j}$ and $R'_{t-5,j}$, and the proportion of fish that return at ages, g = 3, 4, and 5 ($p_{3,t-3,j}$, $p_{4,t-4,j}$, and $p_{5,t-5,j}$, respectively) as:

(F1) $R_{t,j} = R'_{t-3,j} \cdot p_{3,t-3,j} + R'_{t-4,j} \cdot p_{4,t-4,j} + R'_{t-5,j} \cdot p_{5,t-5,j}$

The proportion of the mature fish that return at each age, g, including multivariate logistic stochastic variation was calculated as:

(F2) $p_{g,y,j} = \frac{\bar{p}_g e^{\varpi \cdot \varepsilon_{g,y,j}}}{\sum_3^5 \bar{p}_g \cdot e^{\varpi \cdot \varepsilon_{g,y,j}}}$,

where y is the brood year and y= t-3, t-4, or t-5 in Eqn. F1, the summation in the denominator is over ages 3 to 5 (minimum and maximum ages at recruitment), $\bar{p}_g$ is the mean proportion of adult fish that return at age g, ϖ is a parameter that controls interannual variability in proportions of fish returning at each age (set to 0.8), and ε are standard normal deviates (adapted form Holt and Bradford 2011).

Chum Salmon are vulnerable to fishing only in their final year of life during their return migration to freshwater. The number of fish that escape the fishery (escapement, $E_{t,j}$ in return year t and subpopulation j) and spawn was calculated as:

(F3) $E_{t,j} = R_{t,j} \cdot (1 - h_t)$,

where h_t is harvest rate in return year, t. The number of spawners in each subpopulation was calculated from the escapement multiplied by the dispersal matrix, $\boldsymbol{D}$, which accounts for straying among subpopulations:

(F4) $[S_{t,j=1} \quad \cdots \quad S_{t,j=J}] = \boldsymbol{D} \times \begin{bmatrix} E_{t,j=1} \\ \vdots \\ E_{t,j=J} \end{bmatrix}$,

where $\boldsymbol{D}$ is a symmetric matrix with elements that represent the proportion of escapement that stray from subpopulation n (row n) to subpopulation m (column m), and J is the total number of subpopulations (Peacock and Holt 2012).

(E5) $\boldsymbol{D}_{J \times J} = \begin{bmatrix} 1-d & d\tau_{n=1,m=2} & \cdots & d\tau_{n=1,m=J} \\ d\tau_{n=2,m=1} & 1-d & \cdots & d\tau_{n=2,m=J} \\ \vdots & \vdots & \ddots & \vdots \\ d\tau_{n=J,m=1} & d\tau_{n=J,m=2} & \cdots & 1-d \end{bmatrix}$,

d is the stray rate, and $\tau_{n,m}$ is the probability of dispersal from site n to site m, calculated as,

(F6) $\tau_{n,m} = \frac{1}{2\pi\theta^2} exp\left[-\left(\frac{r_{n,m}}{\theta}\right)\right]$.

θ is a dispersion parameter equal to the average dispersal distance, and *r* is the distance between subpopulation *n* and *m*, calculated as the distance between sites situated along a single river that are randomly selected from uniform distribution between 0 (ocean entry) and 100km (head waters) (Peacock and Holt 2012). Values for $\tau_{n,m}$ were normalized so that the row sum of probabilities for dispersal of salmon from site *n* in matrix ***D*** was equal to one (i.e., no fish strayed outside the CU).

The abundance of recruits that spawned in brood year *y*, (where *y=t-3, t-4*, or *t-5* in Eqn. F1), $R'_{y,j}$, was calculated from a Ricker spawner-recruitment relationship for each subpopulation, *j*, including autocorrelated deviations and log-normal stochastic error:

(F7) $R'_{y,j} = S_{y,j}e^{\left(a_j\left(1-\frac{S_{y,j}}{b_j}\right)+\varphi_{y,j}\right)}, \varphi_{y,j} = \rho\varphi_{y-1,j} + v_{y,j}, v \sim MVN\left(-\frac{1}{2}\sigma_v^2, \Sigma_{j\times j}\right),$

$$\Sigma_{j\times j} = \begin{bmatrix} \sigma_v^2 & \rho\sigma_v^2 & \cdots & \rho\sigma_v^2 \\ \rho\sigma_v^2 & \sigma_v^2 & & \rho\sigma_v^2 \\ \vdots & & \ddots & \vdots \\ \rho\sigma_v^2 & \rho\sigma_v^2 & \cdots & \sigma_v^2 \end{bmatrix}_{j\times j}$$

where $a_j = \log(\alpha_j)$, and α_j is recruits/spawner at low spawner abundances for subpopulation *j* (defined as productivity here)*, b_j is the equilibrium abundance in the absence of fishing (S_{eq}), S_{max}, spawner abundances that maximize recruitment, is equal to b/a, φ and v are stochastic terms, ρ is an autocorrelation coefficient, v is drawn from a multivariate normal distribution with mean $-\frac{1}{2}\sigma_v^2$ to make the arithmetic mean of the lognormally distributed recruitment variation equal to 1 (as in Holt and Peterman 2008), Σ is a covariance matrix describing how recruitment residuals covary among subpopulations, and σ_v is the standard deviation in residual error without autocorrelation (Ricker 1975; Holt and Bradford 2011). σ_v is related to the standard deviation in residual error that includes autocorrelated variability, σ_ϵ, through the relationship:

(F8) $\sigma_\epsilon^2 = \sigma_v^2/(1 - \rho^2)$

Given parameters for the base case, σ_v =0.75, and ρ =0.6 derived from the literature (as described in the text), σ_ϵ was calculated as 0.94 from Eqn. F8. The value of σ_ϵ was held constant (and σ_v varied) under sensitivity analyses where ρ changed. Spawner abundances in brood years y=1 to 5 were initialized at 20% of S_{eq} to seed Eqn. F7. That initial abundance level is within the range of variability observed historically for Chum Salmon on the inner south coast of BC (Ryall et al. 1999; Hilborn et al. 2013), and preliminary analyses showed that performance metrics were insensitive to changes in initial abundances from 10%-30% of S_{eq}. $\varphi_{y=1,j}$ was initialized at zero in year 1. $R'_{y=t-3,j}$, $R'_{y=t-4,j}$, and $R'_{y=t-5,j}$ were then applied to Eqn. F1 starting in year 6 to calculate abundance of adult recruits aligned by return year $R_{t,j}$.

OBSERVATION SUB-MODEL

Observed spawner abundances, $\hat{S}_{y,j}$, were estimated from "true" spawner abundances, $S_{y,j}$, in brood year *y*, with log-normal error:

(F9) $\hat{S}_{y,j} = S_{y,j} \cdot e^{\delta_{y,j}}$, $\delta \sim N\left(-\frac{1}{2}\sigma_\delta^2, \sigma_\delta^2\right)$,

where δ is a stochastic term with mean equal to $-\frac{1}{2}\sigma_\delta^2$ to make the arithmetic mean of the lognormally distributed spawner abundances equal to 1, and σ_δ is the standard deviation in residual observation error in spawner abundances. Aggregate observed spawner abundances in brood year y, $\hat{S}_{AG,y}$, were calculated by summing over observed spawner abundances for each subpopulation:

(F10) $\hat{S}_{AG,y} = \sum_{j=1}^{J}\hat{S}_{y,j}$

Observed total catch, $\hat{C}_t$, was estimated from the "true" catch in return year t, C_t, with log-normal error:

(F11) $\hat{C}_t = C_t e^{\chi_t}$, $\chi \sim N\left(-\frac{1}{2}\sigma_\chi^2, \sigma_\chi^2\right)$,

where χ is a stochastic term with mean equal to $-\frac{1}{2}\sigma_\chi^2$ to make the arithmetic mean of the lognormally distributed catches equal to 1, and σ_χ is the standard deviation in residual observation error in catches.

The observed number of returning adult fish in year t was calculated as catch plus spawning escapement in that year:

(F12) $\hat{R}_t = \hat{C}_t + \hat{S}_{AG,y}$

To estimate the recruitment aligned by brood year in year, y-6, from observed data (when the "true" underlying recruitment is unknown), estimates of the proportion of age-3 fish that returned in year t-3, $\widehat{pr}_{g=3,t-3}$, the proportion of age-4 fish that returned in year t-2, $\widehat{pr}_{g=4,t-2}$, and the proportion of age-5 fish that returned in year t-1, $\widehat{pr}_{g=5,t-1}$ are required. Estimated proportions at age were observed from the true proportions with multivariate logistically distributed error,

(F13) $\widehat{pr}_{g,t} = \dfrac{\overline{pr}_{g,t}e^{\varpi_{pr}\cdot\varepsilon_{g,t}}}{\sum_3^5 \overline{pr}_{g,t}\cdot e^{\varpi_{pr}\cdot\varepsilon_{g,t}}}$,

where t is the return year, the summation in the denominator is over ages 3 to 5 (minimum and maximum ages at recruitment), $\overline{pr}_{g,t}$, is the actual proportion of age g fish in the returns, ϖ_{pr} is a parameter that controls interannual variability in proportions of fish returning at each age (set to 0.1, less than the natural variability in age-at-maturity in Eqn. F2), and ε are standard normal deviates. Note, the proportions at age among adult returns $pr_{g,t}$ (Eqn. F13) differ from the proportions of the spawning brood year that returned at each age, $p_{g,y}$ (Eqn. F2).

Observed recruitment $\widehat{R}'_{y-6}$ from brood year y-6 for each subpopulation j was calculated from observed returns in return years $t=y$-3, y-2, and y-1, and the observed proportions at age in those return years.

(F14) $\widehat{R}'_{y-6} = \hat{R}_{t=y-3,}\cdot\widehat{pr}_{g=3,t=y-3} + \hat{R}_{t=y-2}\cdot\widehat{pr}_{g=4,t=y-2} + \hat{R}_{t=y-1}\cdot\widehat{pr}_{g=5,t=y-1}$

MANAGEMENT SUB-MODEL

Four benchmarks were estimated annually in the management sub-model, two lower benchmarks: S_{gen}, the spawner abundances that will generate the level of recruitment required to achieve S_{MSY} in one generation in the absence of fishing and stochastic variation, and S_{25th}, the 25th percentile of observed spawner abundances, and two upper benchmarks: 80% of S_{MSY}, spawner abundances at maximum sustainable yield, and S_{75th}, the 75th percentile of observed

spawner abundances. For the Ricker benchmarks, S_{gen} and 80% of S_{MSY}, the parameters of the standard Ricker spawner-recruitment relationship, $\hat{a}$ and $\hat{b}$, were estimated annually without autocorrelation, using observed spawner, $\hat{S}_{AG,y}$, and recruitment, $\widehat{R}'_y$ data aligned by the year in which they spawned prior to the year of the assessment including initialization period, aggregated over subpopulations:

(F15) $\widehat{R}'_y = \hat{S}_{AG,y} e^{\left(\hat{a}\left(1 - \frac{\hat{S}_{AG,y}}{\hat{b}}\right)\right)}$

Eqn. F15 was reformulated to:

(F16) $log_e\left(\widehat{R}'_y \Big/ \hat{S}_{AG,y}\right) = \hat{a} - \frac{\hat{a}}{\hat{b}}\hat{S}_{AG,y}$

and least squares regression was used to estimate parameters, $\hat{a}$ and $\hat{b}$. S_{MSY} was calculated from the Ricker parameters using an approximation described by Hilborn and Walters (1992):

(F17) $S_{MSY} = \hat{b}(0.5 - 0.07\hat{a})$

S_{gen}, was estimated by setting the spawner abundances, $\hat{S}_{AG}$, to S_{gen} and the recruitment value associated with S_{gen} , $\widehat{R}'$, to S_{MSY} in Eqn. F17:

(F18) $S_{MSY} = S_{gen} \cdot e^{\left(\hat{a}\left(1 - S_{gen}\big/\hat{b}\right)\right)}$,

In Eqn. F18, S_{gen} was solved numerically with constraints between 0 and S_{eq}, and S_{MSY} was calculated from Eqn. F17 above (Holt and Bradford 2011). The percentile benchmarks, S_{25th} and S_{75th}, were calculated from the 25th and 75th percentiles of observed aggregate spawner abundance when ordered from lowest to highest abundances, using all data prior to the year of assessment including the initialization period. In sensitivity analyses, alternative percentiles were considered.

HARVEST SUB-MODEL

To model outcome uncertainty, annual harvest rates were calculated from the target harvest rate, h'_t, with normally distributed stochastic error, ω_t,

(F19) $h_t = h'_t + \omega_t, \omega_t \sim N(0, \sigma_\omega^2)$,

where σ_ω is the standard deviation in residual error (Holt and Bradford 2011). Harvest rates were constrained to be between 0 and 1, where values >1 and <0 were resampled. Catches were calculated from the realized harvest rate and total adult returns in that year, aggregated over sub-populations.

(F20) $C_t = h_t \cdot \sum_{j=1}^{J} R_{t,j}$.

COMPARISON WITH HOLT AND FOLKES (2015)

Our simulation model was adapted from that described by Holt and Folkes (2015), but differed in several ways. Most importantly,

- The population dynamics sub-model included covariance in Ricker residuals among subpopulations within a CU, instead of assuming sub-populations varied independently.

- A back-transformation bias correction for log-normally distributed variables (e.g., residual variation from Ricker model, observation error in spawner abundances) was included to

make the arithmetic mean of the lognormally distributed recruitment variation equal to 1 (as in Holt and Peterman 2008).

- The observation sub-model was more realistic in that catches were observed with observation errors, and recruitment by brood year was 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..

- 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.

- The assessment sub-model model focused on percentile-based benchmarks and stock-recruitment benchmarks applied under the Wild Salmon Policy (S_{gen} and 80% of S_{MSY} for the lower and upper benchmarks, respectively).

- In the harvest sub-model, a constant low harvest rate was applied instead of a harvest control rule with limit and/or target management reference points.

- In the performance module, benchmarks were evaluated based on the deviations between benchmark estimates and the "true" underlying values.

- The model was run over 50 years, instead of 100 to provide a more realistic time-series length for estimating benchmarks.

APPENDIX F REFERENCES

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, Public Certification Report. Intertek Moody Marine and the Canadian Pacific Sustainable Fisheries Society. (Accessed February 18, 2018)

Hilborn, R.; Walters, C.J. 1992. Quantitative Fisheries Stock Assessment. New York: Chapman and Hall.

Holt, C.A.; Bradford, M.J. 2011. Evaluating benchmarks of population status for Pacific salmon. North Am J Fish Manag. 31, 363-378.

Holt C.A.; Folkes M.J.P. 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 C.A.; Peterman R.M. 2008. Uncertainties in population dynamics and outcomes of regulations in sockeye salmon (*Oncorhynchus nerka*) fisheries: implications for management. Can. J. Fish. Aquat. Sci. 65:1459-1474.

Peacock, S.J.; Holt, C.A. 2012. Metrics and sampling designs for detecting trends in the distribution of spawning Pacific salmon (*Oncorhynchus* spp.). Can. J. Fish. Aquat. Sci. 69, 681-694.

Ricker, W.E., 1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Board Can. 191, 382 p.

Ryall, P.; Murray, C.; Palermo, V.; Bailey, D; Chen, D. 1999. Status of Clockwork Chum Salmon Stock and Review of the Clockwork Management Strategy. DFO Can. Sci. Advis. Sec. Res. Doc. 1999/169. 134 p.

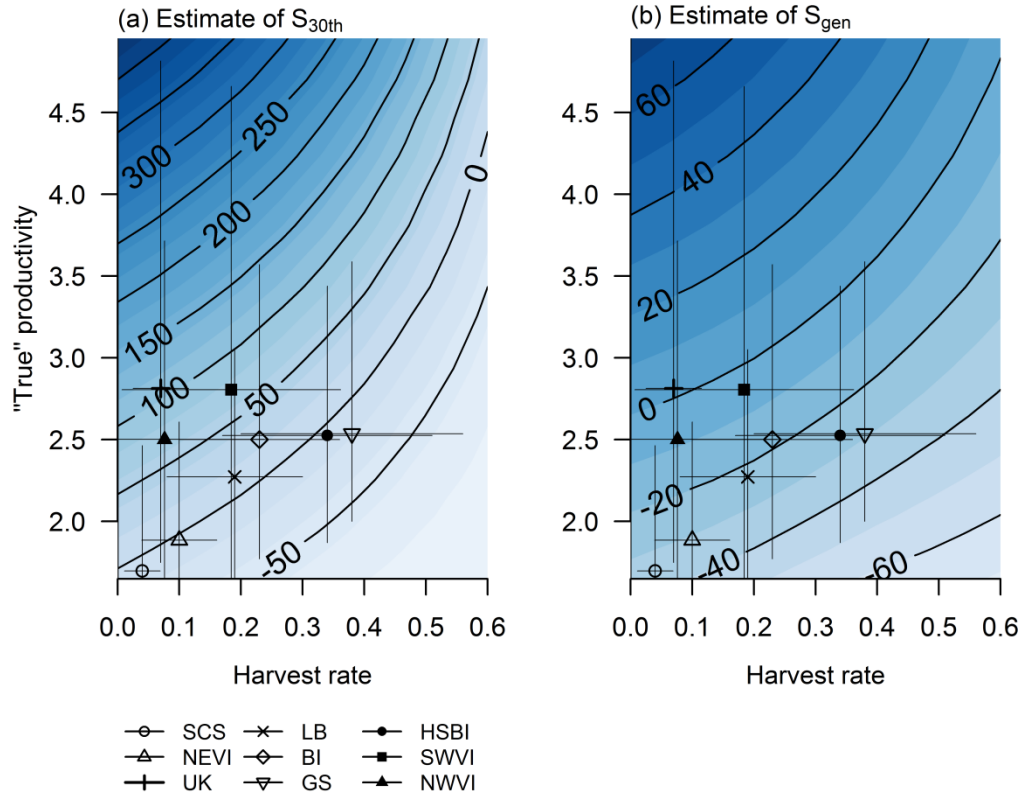# APPENDIX G. ALTERNATIVE PERCENTILE-BASED LOWER BENCHMARKS

Figure G1. Mean percent error, MPE, of the estimated lower benchmark, S30th (a), and Sgen (b) from the "true" lower benchmark ("true" Sgen) along a gradient in harvest rates and "true" productivities derived from a simulation model of a hypothetical Chum Salmon CU. Symbols indicate CU-specific productivities and harvest rates. 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 Northeast Vancouver Island, UK is Upper Knight, LB is Loughborough, GS is Georgia Strait, and HSBI is Howe Sound/Burrard Inlet, NWVI is Northwest Vancouver Island, and SWVI is Southwest Vancouver Island.
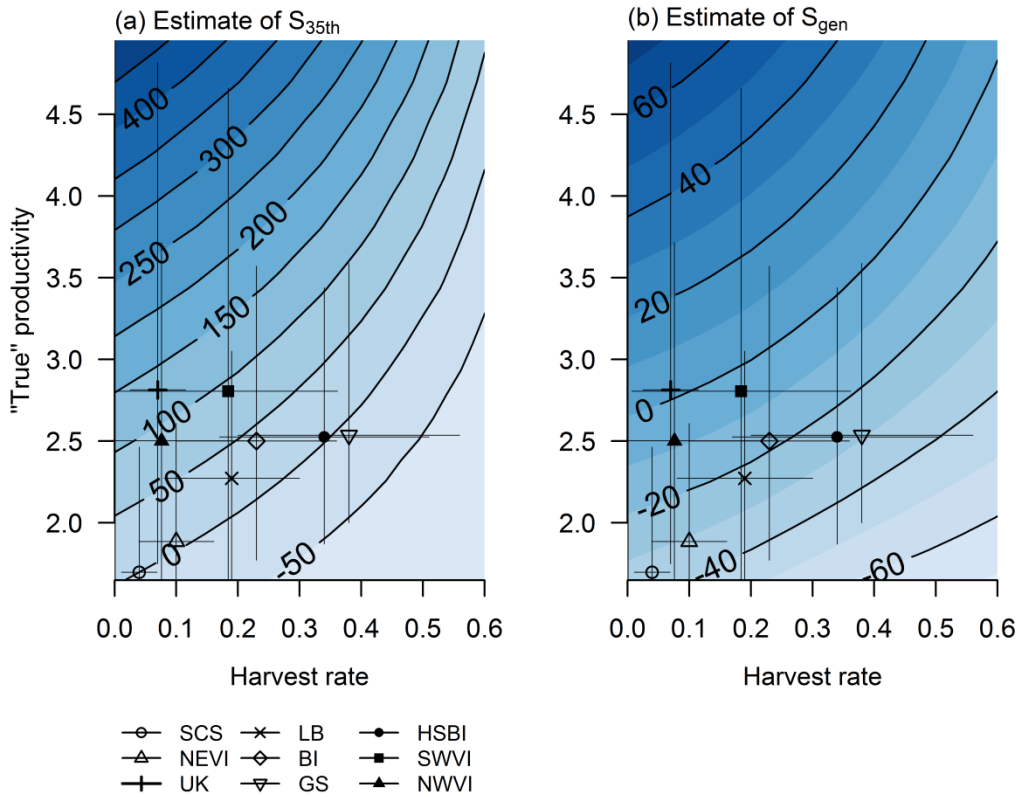
Figure G2. Mean percent error, MPE, of the estimated lower benchmark, S_{35th} (a), and S_{gen} (b) from the "true" lower benchmark ("true" S_{gen}) along a gradient in harvest rates and "true" productivities derived from a simulation model of a hypothetical Chum Salmon CU. Symbols and lines are described in the caption to Fig. G1.Note (b) is the same as in Fig. G1, but is shown here to facilitate comparison with panel (a).
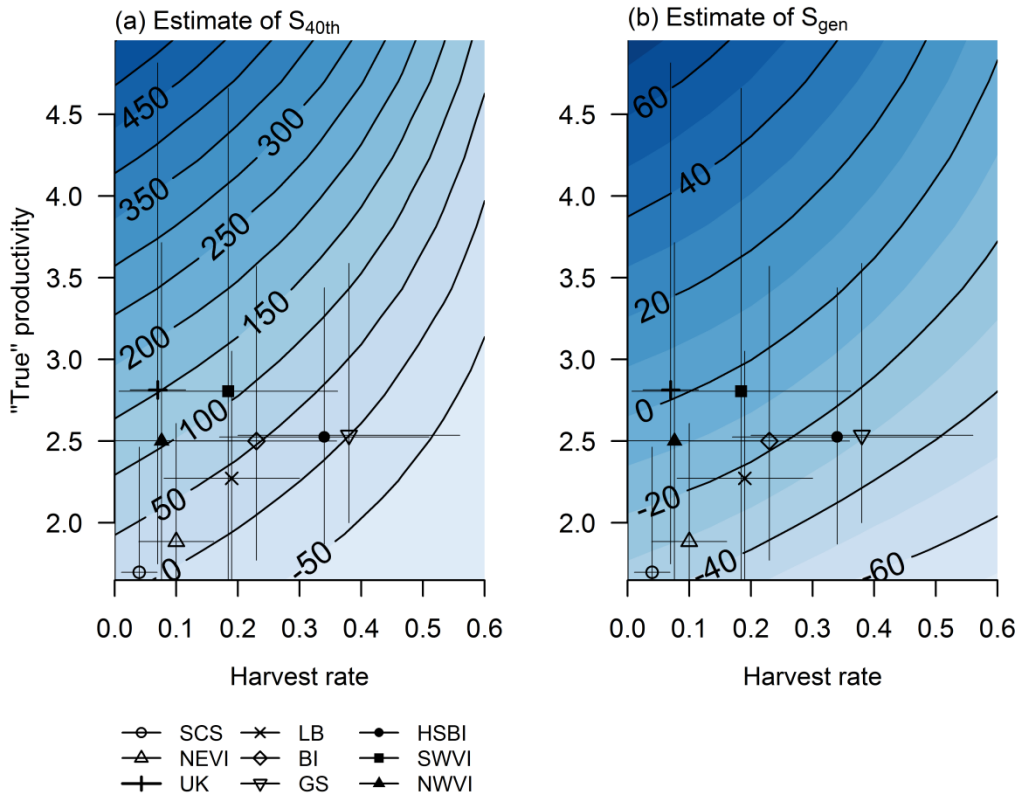
Figure G3. Mean percent error, MPE, of the estimated lower benchmark, S_{40th} (a), and S_{gen} (b) from the "true" lower benchmark ("true" S_{gen}) along a gradient in harvest rates and "true" productivities derived from a simulation model of a hypothetical Chum Salmon CU. Symbols and lines are described in the caption to Fig. G1. Note (b) is the same as in Fig. G1, but is shown here to facilitate comparison with panel (a).
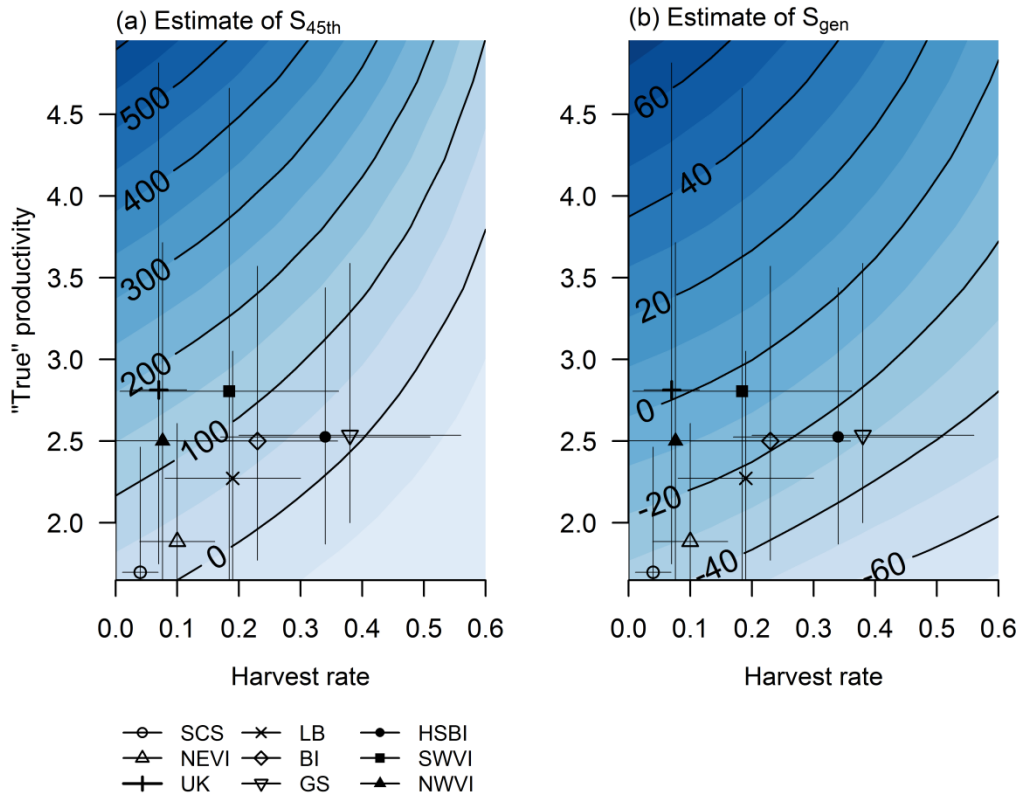
Figure G4. Mean percent error, MPE, of the estimated lower benchmark, S_{45th} (a), and S_{gen} (b) from the "true" lower benchmark ("true" S_{gen}) along a gradient in harvest rates and "true" productivities derived from a simulation model of a hypothetical Chum Salmon CU. Symbols and lines are described in the caption to Fig. G1. Note (b) is the same as in Fig. G1, but is shown here to facilitate comparison with panel (a).
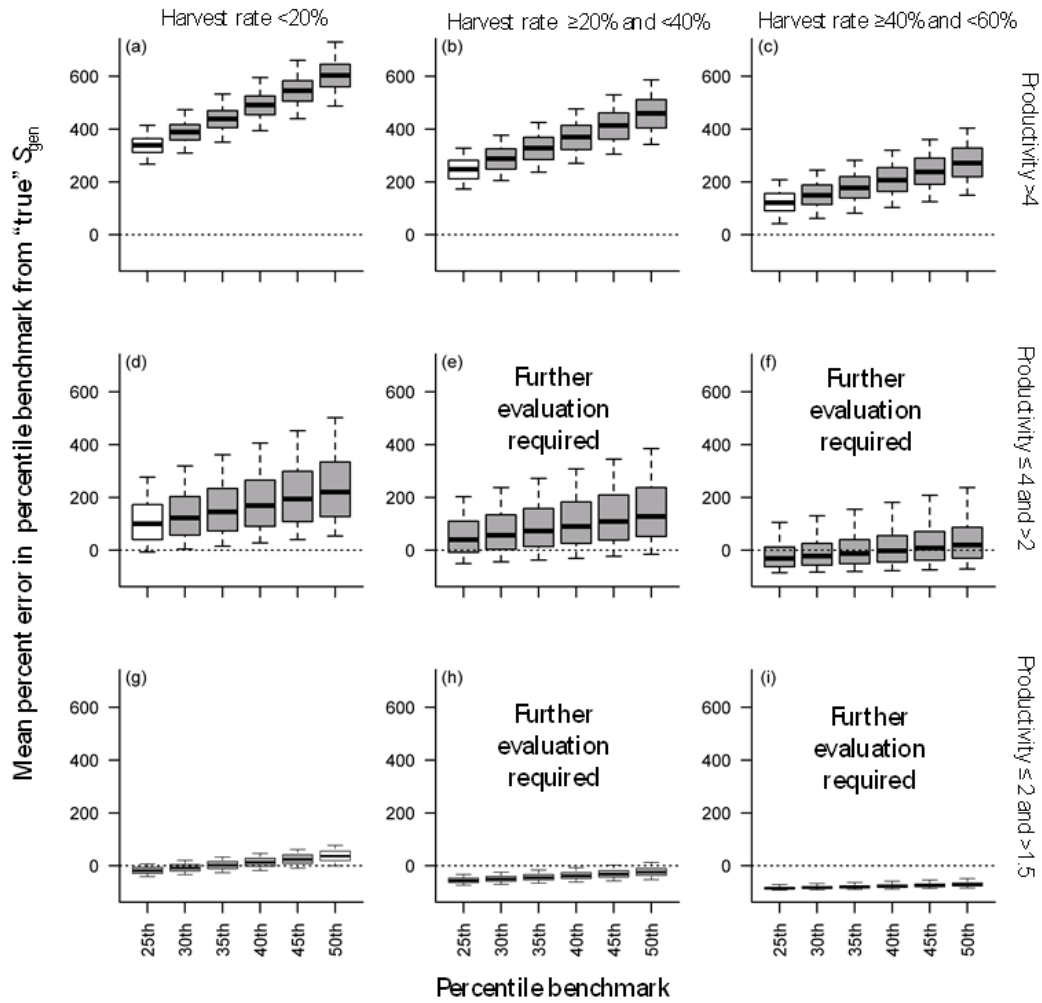
Figure G5. Distributions of mean percent error, MPE, of various percentile lower benchmarks (labelled on x-axis) from "true" S$_{gen}$ along a gradient in harvest rates (columns, labelled at top) and productivities (rows labeled on right), derived from a simulation model of a hypothetical Chum Salmon CU. Distributions for each benchmark are derived from simulated outputs in Fig.13, Fig. 14 and Figs. G1-G4, aggregated within productivity and harvest rate groupings. Groupings for productivity differ from those in Fig.15.
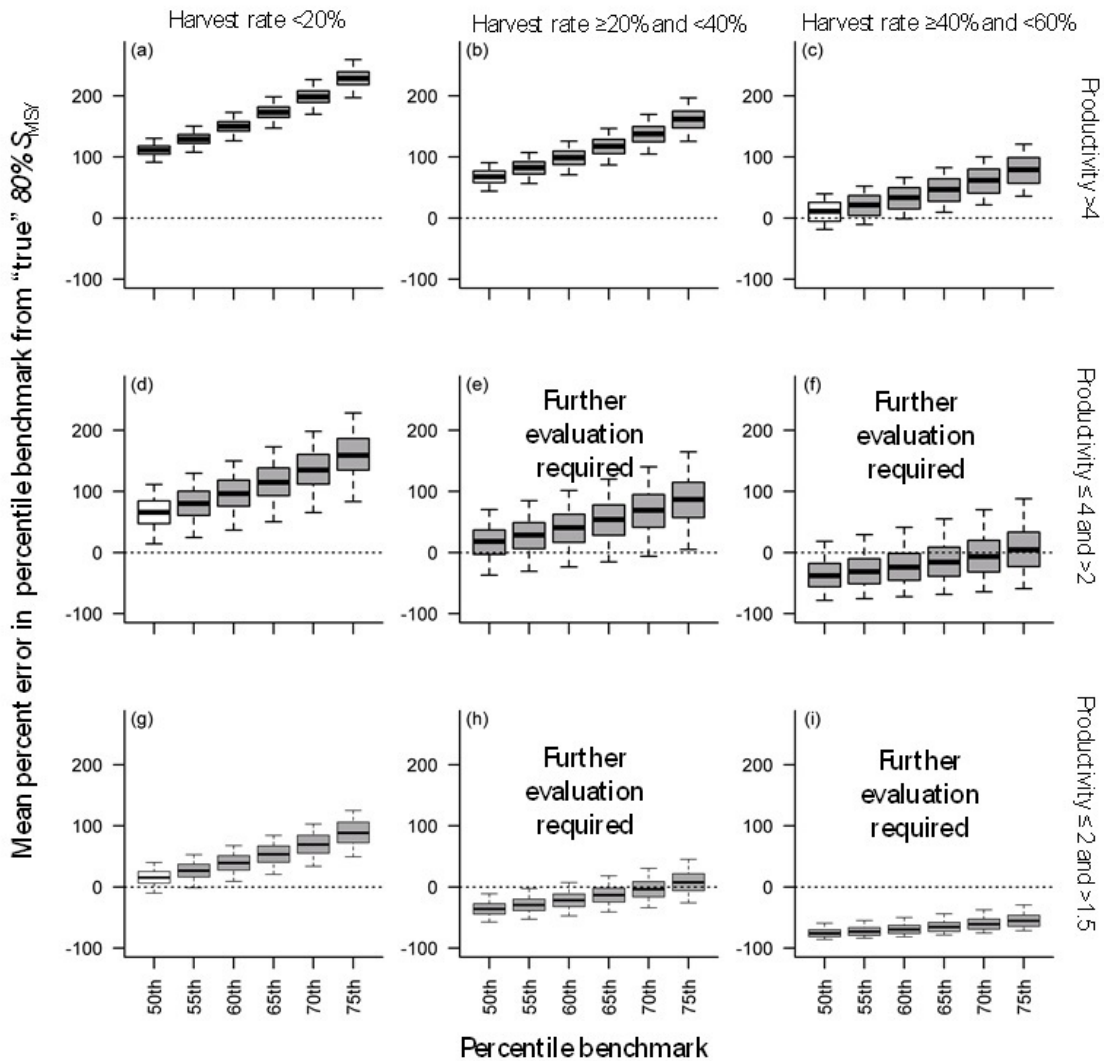
Figure G6. Distributions of mean percent error, MPE, of various percentile upper benchmarks (labelled on x-axis) from "true" 80% S_{MSY} along a gradient in harvest rates (columns, labelled at top) and productivities (rows labeled on right), derived from a simulation model of a hypothetical Chum Salmon CU. Distributions for each benchmark are derived from simulated outputs aggregated within productivity and harvest rate groupings. Groupings for productivity differ from those in Fig.19.

APPENDIX H. GLOSSARY

Biological Benchmark: Values of quantitative metric that delineate red, amber, and green zones of biological status under the Wild Salmon Policy, defined based on biological criteria

Capacity: Spawner abundances that will maximize recruitment according to a Ricker stock-recruitment model.

Conservation Unit: The unit of conservation under Canada's Wild Salmon Policy. It is defined as a group of wild salmon sufficiently isolated from other groups that, if extirpated is very unlikely to recolonize naturally within an acceptable timeframe, such as a human lifetime or a specified number of salmon generations.

Data-deficient Conservation Unit: A Conservation Unit without time-series of spawning abundances or adult recruitment required to generate assessments under the Wild Salmon Policy.

Data-limited Conservation Unit: a Conservation Unit with time-series of escapement (relative or absolute), but without time-series of adult recruitment required to estimate stock-recruitment based benchmarks on abundances.

Data-rich Conservation Unit: a Conservation Unit with time-series of absolute spawning abundances and adult recruitment, for which stock-recruitment based benchmarks can be estimated under the Wild Salmon Policy.

Hierarchical Model: models that contain more than one level of organization, such as multiple CUs within a region, for at least one parameter (e.g., the productivity parameter in the example in the current study). In this example, CU-specific productivity parameters are drawn from a distribution of regional distribution.

Integrated Status Assessment: Status assessments under the Wild Salmon Policy that combine information across numerous metrics and benchmarks (e.g., relative abundances, short-term trends, long-term trends). Examples include Grant and Pestal (2012) and DFO (2016) for Fraser River Sockeye Salmon and Southern BC Chinook Salmon, respectively.

Management Reference Point: Quantitative states that characterize desirable biological and/or economic properties of fisheries

Productivity: Recruits per spawner at low spawner abundances estimated from a stock-recruitment model.