

# Evaluation of performance of alternative management models of Pacific salmon (*Oncorhynchus* spp.) in the presence of climatic change and outcome uncertainty using Monte Carlo simulations

Brigitte Dorner, Randall M. Peterman, and Zhenming Su

**Abstract:** An important management challenge is to maintain productive populations of Pacific salmon (*Oncorhynchus* spp.), despite highly variable environments and our weak understanding of future climatic conditions and mechanisms that link them to salmon. This understanding could be improved by including environmental covariates in salmon population models and applying advanced “meta-analyses” to large data sets to better estimate underlying functional relationships. However, the performance of such models needs to be determined in the context of an overall system. We therefore simulated a 15-population salmon fishery system and compared the performance (in terms of catch and an index of conservation concern) of 10 forecasting and stock assessment models, ranging from simple to complex, by stochastically simulating components of a salmon fishery using a “closed-loop simulation” (or “management strategy evaluation”) under a variety of plausible future climatic scenarios. We found that complex models perform better in some situations. However, their incremental benefits are small and are swamped by the large variability in outcomes of management actions caused by “outcome uncertainty”, which reflects noncompliance of fishing vessels with regulations as well as variation in catchability. Reduction of this outcome uncertainty should therefore be a top priority, as should evaluations of more complex stock assessment models before adopting them.

**Résumé :** C'est un défi important de gestion que de maintenir des populations très productives de saumons du Pacifique (*Oncorhynchus* spp.) malgré les environnements très variables et une faible compréhension des conditions climatiques futures et des mécanismes qui les relient aux saumons. Cette compréhension pourrait être améliorée si on ajoutait des covariables environnementales dans les modèles démographiques de saumons et si on faisait des « méta-analyses » des grands ensembles de données afin de mieux estimer les relations fonctionnelles sous-jacentes. Cependant, la performance de tels modèles doit être déterminée dans le contexte d'un système global. Nous avons donc simulé un système de pêche de saumons comprenant 15 populations et comparé la performance (en ce qui a trait à la capture et à un indice de préoccupation pour la conservation) de 10 modèles de prévision et d'évaluation des stocks, allant de simples à complexes. Nous simulons de façon stochastique les composantes d'une pêche à saumons à l'aide d'une « simulation en boucle fermée » (ou d'une « évaluation des stratégies de gestion ») dans une variété de scénarios climatiques plausibles dans le futur. Les modèles complexes fonctionnent mieux dans certaines circonstances. Cependant, leur bénéfice additionnel est petit et masqué par l'importante variabilité des résultats des actions de gestion causée par « l'incertitude des résultats » qui reflète le fait que les navires de pêche ne se conforment pas aux réglementations et que la capturabilité est variable. On devrait donc donner une forte priorité à la réduction de cette « incertitude des résultats »; il en va de même des évaluations des modèles plus complexes d'estimation des stocks avant leur adoption.

[Traduit par la Rédaction]

## Introduction

Scientists and fisheries managers who work on Pacific salmon (*Oncorhynchus* spp.) face major challenges. They seek

to maintain abundant and productive populations in spite of large uncertainties in four areas. (i) Freshwater and marine survival rates of Pacific salmon are highly variable over space and time, which makes forecasting, abundance estima-

Received 2 December 2008. Accepted 20 August 2009. Published on the NRC Research Press Web site at cjfas.nrc.ca on 20 November 2009.

J20901

Paper handled by Associate Editor Ray Hilborn.

**B. Dorner,<sup>1</sup> R.M. Peterman,<sup>2</sup> and Z. Su.<sup>3</sup>** School of Resource and Environmental Management, 8888 University Drive, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

<sup>1</sup>Present address: Driftwood Cove Designs, GD Lasqueti Island, BC V0R 2J0, Canada.

<sup>2</sup>Corresponding author (e-mail: peterman@sfu.ca).

<sup>3</sup>Present address: Institute for Fisheries Research, Michigan Department of Natural Resources and University of Michigan, 1109 North University Avenue, Ann Arbor, MI 48109-1084, USA.

tion, and choices of management options difficult. (ii) Details of mechanisms that link salmon survival rates to environmental conditions are still not yet clearly understood, despite considerable progress over the last two decades at identifying correlates with salmon survival rates (e.g., Beamish 1995). (iii) Even if those mechanisms become clear, future regional climatic conditions that drive environmental variables, let alone their effects on salmon, are still uncertain, except in the most general terms. “Downscaling” those forecasts to smaller spatial regional responses in aquatic systems adds another layer of uncertainty (Intergovernmental Panel on Climate Change 2007). (iv) The effectiveness of proposed management actions is clouded by outcome uncertainty (also referred to as implementation error, e.g., Rosenberg and Brault 1993), which causes realized spawning escapements or harvest rates to differ from managers’ targets. These deviations from management targets occur for at least three reasons. Catchability can vary considerably and unpredictably due to biological or physical environmental conditions that affect vulnerability of fish to fishing gear. As well, preseason and in-season forecasts are imperfect, and there is a time delay in obtaining catch and escapement information and feeding it into in-season updates for managers to use. Also, lack of compliance by harvesters and (or) weak enforcement of regulations can lead to deviations from management targets. This resulting outcome uncertainty has been shown to have substantial effects on outcomes of management options in some cases (Cass et al. 2003; Kell et al. 2005).

One way to deal more effectively with at least some sources of uncertainties has emerged from the recent development of advanced parameter estimation techniques and compilations of large fisheries databases. The late Ransom A. Myers was the main proponent of using hierarchical statistical models in such situations to conduct “meta-analyses” to answer various questions using large databases of multiple fish populations (reviewed in Myers 2001). Such hierarchical, multistock models simultaneously estimate model parameters for all stocks of a given species within a specified area rather than conducting separate analyses stock by stock. In our case of Pacific salmon, this hierarchical method is justified because productivities of multiple stocks of sockeye (*Oncorhynchus nerka*), pink (*Oncorhynchus gorbuscha*), and chum salmon (*Oncorhynchus keta*) do not all vary independently (Dorner et al. 2008); instead, some of them within regions respond similarly to shared environmental drivers (Peterman et al. 1998; Pyper et al. 2001, 2002). In such situations, multistock hierarchical models help to average out unavoidable observation errors that create spurious variation in estimates of abundance and productivity (Thompson and Page 1989). Hierarchical models essentially “borrow strength”, or utilize information, from other correlated stocks to improve estimates of parameters for individual stocks (Banerjee et al. 2004). For instance, for a database of 120 Pacific salmon populations, several analyses with hierarchical models have shown better fits to historical spawner–recruit data than fitting the analogous single-stock model separately to each population (Mueter et al. 2002a; Su et al. 2004; Peterman et al. 2009). Parameter estimates from hierarchical models of effects of sea surface temperature on productivity of those salmon populations also gave more precise values than in the single-stock cases.

Although better fits to the entire set of historical data are obtained by using hierarchical models rather than analyzing each population separately, the predictive performance of hierarchical models under different future climatic conditions and management regimes needs to be tested. To fully understand the benefits and drawbacks of these and other methods used in stock assessment and forecasting models, they should be explored across hundreds of simulated possible future environmental conditions and other sources of uncertainty.

Therefore, the purpose of our research was to evaluate the relative performance of several combinations of harvest policies and stock assessment/forecasting models, including hierarchical models and models that make use of environmental covariates, in the presence of uncertainties about future climatic conditions and outcome uncertainty. By “evaluate”, we mean determine which combination might best meet management objectives related to harvests and conservation risks.

For these evaluations, we used closed-loop Monte Carlo simulations (Hilborn 1979; Walters 1986), more recently called management strategy evaluations (MSEs) (de la Mare 1998; Sainsbury 1998; Smith et al. 1999) and the management procedure approach (Butterworth 2007). By developing such Monte Carlo simulation models that reflect major uncertainties, fisheries scientists have made significant progress in dealing with the sources of uncertainty about environmental conditions, biological responses to environmental drivers, and outcomes of management decisions. These models go well beyond simple stochastic simulations; they simulate major feedback linkages in an entire fishery system. These linkages include (i) the hypothesized “true” stochastic environmental dynamics and fish populations, (ii) the data collection process (to reflect sources of sampling error and imperfect information), (iii) analysis of those data with a stock assessment model (to reflect the parameter estimation process), (iv) a state-dependent method indicating how regulatory decisions are made based on that assessment, and (v) deviations between actual outcomes and target harvest rates or spawning goals (to reflect outcome uncertainty). Such models mimic feedbacks within fishery systems more closely than standard types of simulation models that do not attempt to represent an entire fishery system. Usually, the ultimate purpose of these MSEs is to identify combinations of data collection and analysis methods, decision-making processes based on those stock assessments, and harvest guidelines or control rules that are most robust to all of the uncertainties considered, i.e., the ones that best meet management objectives. Although such MSEs have been conducted for many groundfish, pelagic, and other marine species, they have only rarely been done for salmonids (Walters 1986; Peterman et al. 2000; Holt and Peterman 2008).

Our research is also appropriate for a broader reason. The tendency of stock assessment scientists to develop ever more complex models needs to be justified by showing that such models improve the ability to meet management objectives. Uncertainties about structural forms of stock assessment models are only one source of uncertainty in fishery systems; there are also observation errors in data, inconsistent decision-making procedures, and unexpected responses by harvesters to management regulations, among others. The latter three types of uncertainty might tend to reduce, or even overwhelm, the benefits of an improved stock assessment

model or some new parameter estimation method. Closed-loop MSEs are currently the best way to evaluate the net effect of the aggregate of such sources of uncertainty.

## Materials and methods

### Data

Prior to discussing the simulation model's components, we first describe the data used to estimate its parameters. Our intent in using these data sets was not to simulate any particular stock or region in detail, but rather to generate realistic generic scenarios that represent typical or average conditions experienced by North American sockeye salmon. Our main database consisted of time series of spawner and recruit abundances, the latter stratified by age at return, for 37 stocks of wild North American sockeye salmon with spawning areas ranging from northern Washington to western Alaska and covering brood years from the 1950s to the late 1990s (Table 1) (also see Pyper et al. 2005). In some cases, we parameterized alternative scenarios based on data from distinct spatial aggregates of sockeye salmon stocks with specific characteristics shared among those stocks. For parameter values that were not of direct interest in this analysis and that therefore stayed the same across scenarios, we generally used average values across all sockeye stocks in our database.

We also used additional, more detailed data sets for Fraser River, British Columbia, and Bristol Bay, Alaska, sockeye salmon to parameterize the outcome uncertainty (OU) module, which generates the deviations of actual escapements from stated escapement goals (Holt and Peterman 2006). Data on OU for the Fraser River were provided by Ian Guthrie (Pacific Salmon Commission, 600-1155 Robson Street, Vancouver, British Columbia, personal communication) and included annual preseason escapement goals and forecasts of recruitment, in-season escapement goals and forecasts, and postseason estimates of actual recruitment and escapement for 1986–2003 for the four major sockeye salmon run-timing groups, Early Stuart, Early Summer, Summer, and Late (also see Holt and Peterman 2006). Data on OU for Bristol Bay were provided by Lowell Fair (Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, Alaska, personal communication) and included preseason escapement goals and forecasts as well as postseason estimates of actual escapement and recruitment for 1962–2004. OU data for Bristol Bay covered the same nine stocks as our main database (Table 1), except OU data were for the entire Nushagak River Basin rather than just the Nuyakuk tributary.

Coastal sea surface temperature anomalies derived from the Comprehensive Ocean-Atmosphere Data Set (COADS) ([www.cdc.noaa.gov/coads/](http://www.cdc.noaa.gov/coads/)) served as a real-world model for our simulated environmental index that was included in some forecasting and stock assessment models to assess potential improvement in performance (see Mueter et al. (2002b) for details on how sea surface temperature anomalies were calculated).

### Closed-loop simulation model, CLIM2

To address the challenge of evaluating a range of stock assessment models in the presence of numerous sources of

uncertainty, we developed a closed-loop simulation model, CLIM2, that simulates multiple sockeye salmon stocks and environmental influences on them. It allows for spatial and temporal covariation in the simulated climate-driven productivity patterns of different salmon stocks, thereby making it possible to evaluate the performance of stock assessment models (such as hierarchical models) that take advantage of empirically documented similarities among salmon stocks in their temporal variation in productivity (Pyper et al. 2005).

CLIM2 represents the components of a full fishery system using the following seven modules (Fig. 1), which are explained in detail later. (1) A salmon population module simulates annual stochastic dynamics of 15 hypothetical (sockeye-salmon-like) populations. (2) The climatic and environment module determines annual productivities of these salmon populations. We examined four alternative future low-frequency climate-induced patterns in productivity over time and three levels of spatial coherence (i.e., spatial autocorrelation or covariation across space) in those patterns. In addition to time series of salmon productivity, the module generates an environmental covariate time series with a user-specified correlation between the environmental index and salmon productivity. We examined three levels of that correlation to determine the relative merits of models that could use that environmental index as an independent variable. (3) The simulated data collection step for spawners and recruits reflects observation error as well as limited data availability. (4) The stock assessment module fits one of seven structural forms of stock assessment models to the simulated observed data to estimate parameters for each salmon stock. (5) The forecasting module uses one of 10 models to estimate annual abundance of adult salmon recruits. (6) Target escapements are set in the management decision module based on annually updated parameter estimates derived from module 4. We also explored three levels of safety margins that were applied to target escapements to reduce conservation concerns. (7) Harvesting of recruits occurs after applying one of three types of OU, which generates differences between target and realized escapements based on forecasted and (or) actual abundance.

To produce results that can be generalized across many situations, CLIM2 was iteratively run across all combinations of the alternative scenarios and model choices indicated above. To our knowledge, these elements have never before been combined in one comprehensive analysis of any fish species, let alone salmon. Other unique aspects of our analysis are simulations of 15 salmon populations and spatial covariation among them (which permits evaluation of multistock hierarchical models) and the large number of qualitatively different forecasting/stock assessment models combined with OU.

CLIM2 was written in C++ and is an extension of the initial CLIM closed-loop simulation model (Peterman et al. 2000). It uses the Gnu GSL library ([www.gnu.org/software/gsl/](http://www.gnu.org/software/gsl/)) for random number generation and most numerical computations. Due to the large number of sets of input conditions (scenarios) that we explored and sensitivity analyses that we conducted, as well as the CPU-intensive Bayesian estimation methods used for the hierarchical models, our simulations were run on two dual-core CPU clusters made available through Simon Fraser University's Academic

**Table 1.** Summary of 37 wild sockeye salmon data sets used to parameterize the model.

Region	Stock	Brood years	<i>N</i>	<i>R</i>	Source
Puget Sound	Lake Washington	1967–1993	27	274	1
Fraser Early Stuart	Early Stuart	1950–1996	47	340	1
Fraser Early Summer run	Bowron	1950–1996	47	45	1
	Gates	1952–1996	45	46	1
	Nadina	1950–1996	47	68	1
	Pitt	1950–1996	47	72	1
	Raft	1950–1996	47	28	1
	Seymour	1950–1996	47	138	1
Fraser Summer run	Chilko	1950–1996	47	1533	1
	Horsefly	1950–1996	47	1322	1
	Late Stuart	1950–1996	47	585	1
	Stellako	1950–1996	47	499	1
Fraser Birkenhead	Birkenhead	1950–1996	47	424	1
Fraser Late run	Adams	1950–1996	47	2026	1
	Cultus	1950–1996	47	67	1
	Portage	1951–1996	43	45	1
	Weaver	1950–1996	47	289	1
BC central coast	Long Lake	1980–1994	15	350	2
BC north coast	Skeena	1950–1994	45	2056	2
	Nass	1967–1991	25	650	3
Prince William Sound	Copper	1961–1993	33	1435	4
Cook Inlet	Cook	1968–1992	25	4770	5
Kodiak	Ayakulik	1965–1993	29	574	6
	Frazer	1965–1993	29	409	6
	Early Upper Station	1969–1993	25	104	6
	Late Upper Station	1970–1993	24	539	6
Chignik	Black	1950–1993	44	1028	6
	Chignik	1950–1993	44	960	6
Bristol Bay	Branch	1956–1995	40	499	7
	Egegik	1956–1995	40	6424	7
	Igushik	1956–1995	40	965	7
	Kvichak	1956–1995	40	12450	7
	Naknek	1956–1995	40	3487	7
	Nuyakuk	1956–1982	27	932	7
	Togiak	1956–1995	40	518	7
	Ugashik	1956–1995	40	2729	7
	Wood	1956–1995	40	2715	7

**Note:** Stocks are arranged from south to north. *N* is the number of brood years available and *R* is the average number of annual recruits (thousands). Source: 1, Jim Woodey and Mike Lapointe, Pacific Salmon Commission, 600-1155 Robson Street, Vancouver, BC V6E 1B5 (personal communication); 2, Chris Wood, Canadian Department of Fisheries and Oceans, 3190 Hammond Bay Road, Nanaimo, BC V9T 6N7 (personal communication); 3, Les Jantz, Canadian Department of Fisheries and Oceans, 417-2nd Avenue West, Prince Rupert, BC V8J 1G8 (personal communication); 4, Mark Willette, Alaska Department of Fish and Game, 43961 Kalifornsky Bch Road, Soldotna, AK 99669-8367, USA (personal communication); 5, Ken Tarbox, Alaska Department of Fish and Game, 43961 Kalifornsky Bch Road, Soldotna, AK 99669-8367, USA (personal communication); 6, Patricia Nelson, Alaska Department of Fish and Game, 211 Mission Road, Kodiak, AK 99615-6399 (personal communication); 7, Michael Link, Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK 99518-1599, USA (personal communication).

Computing Services and WestGrid. The latter is a network of high-performance computing resources shared among western Canadian universities. We ran our analyses on 3.06 GHz dual-core Intel Xenon processors and 2.133 dual-core Athlon MP 2800+ processors ranging in CPU time from several minutes to 2–4 days per scenario, the latter for the hierarchical Bayesian models.

### Model components

The first two modules, salmon population dynamics and climatic and environmental changes, constitute CLIM2's

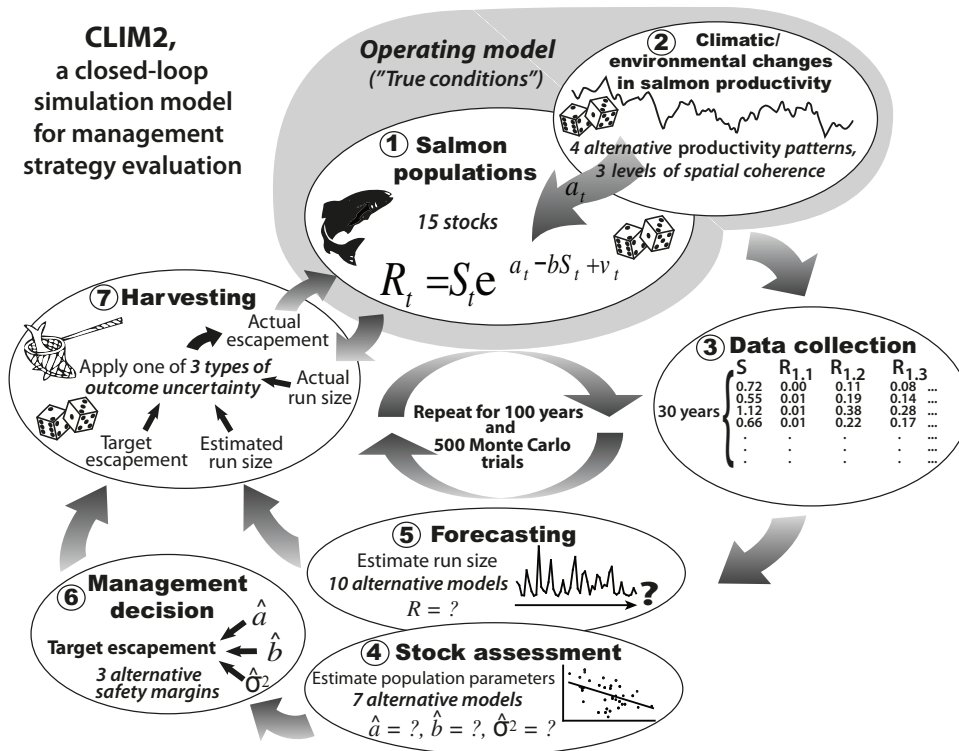
“operating model” (Fig. 1), which is most generally defined as a means to generate simulated data to reflect hypothesized “true” situations (Linhart and Zucchini 1986). Below, modules of CLIM2 are identified by numbers corresponding to Fig. 1.

### Salmon populations

The salmon population component (Fig. 1, ellipse 1) simulates the full life cycle of 15 wild salmon populations from spawning to return to natal rivers. For each population, we used a Ricker spawner–recruit model (Ricker 1975):



**Fig. 1.** Overview of the CLIM2 model, which simulates annual changes in the natural salmon system and its environment (components 1 and 2) and human systems (components 3–7). Each Monte Carlo trial consists of 100 years, and 500 replicate trials were run for each scenario. Dice indicate stochastic model components.



$$(1) \quad \log_e(R_t/S_t) = a_t - bS_t + v_t \text{ or equivalently} \\ R_t = S_t e^{a_t - bS_t + v_t}$$

where  $S_t$  represents the abundance of spawners in brood year  $t$ ,  $R_t$  is abundance of adult recruits from those spawners,  $a_t$  reflects the productivity in brood year  $t$ ,  $b$  reflects density-dependent effects (assumed constant over time for the purpose of this analysis), and  $v_t \sim N(0, \sigma_v^2)$  represents random interannual variability and observation error in recruitment. As in Peterman et al. (2000),  $b$  was set to 1 because it is an arbitrary scalar. Here,  $\sigma_v$  was set to 0.55, which was chosen so that long-run interannual variability of abundance of CLIM2 stocks approximated the overall interannual variability observed in the 37 North American sockeye stocks. To represent different "true" but unknowable underlying models of population dynamics, the  $a_t$ , or productivity, parameter varied across time and among the 15 stocks according to different assumptions in different model runs, as described below.

We simulated full age-structure dynamics so that we would be able to evaluate the sibling model for making pre-season forecasts of adult recruit abundance; this model is widely used for sockeye salmon (Wood et al. 1997). After generating total annual abundance of recruits using eq. 1, we randomly allocated recruits into age groups using the multivariate logistic distribution from Schnute and Richards (1995). To parameterize this distribution so that it would generate historically observed ages, we averaged the natural logarithms of relative proportions of fish in each age stanza first over all past brood years and then over the 37 sockeye stocks to derive target proportions for the logistic distribu-

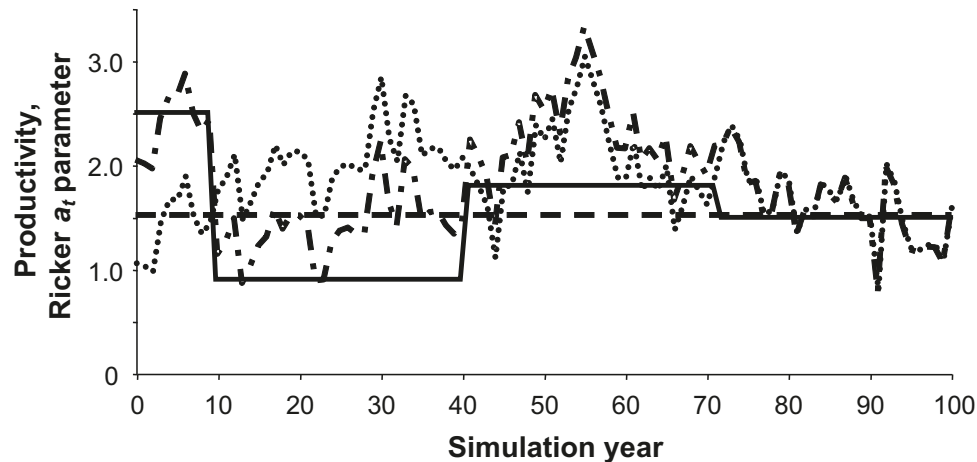
tion. The parameter of the multivariate logistic distribution that controls the interannual variability in the relative proportion of fish returning at each age (the  $\tau$  parameter of Schnute and Richards 1995) was set to 0.9 to emulate the observed temporal variability in age structure of the 37 sockeye stocks.

To initialize the spawning population,  $S_t$ , data were drawn randomly from a uniform distribution between 0.2 times and 0.8 times equilibrium abundance until enough years of data were available to create simulated returns for all adult age groups using eq. 1. For all subsequent simulation years,  $S_t$  was calculated annually by deducting harvested fish from the simulated total adult recruits (returns). In our context, the long-term average value of  $a_t$  is 1.75 and  $b = 1$ , so equilibrium abundance was  $a/b = 1.75$ .

### Climate-induced changes in productivity and environmental covariates

Because future changes in climatic patterns and productivity of salmon are unknown, we evaluated the models across a wide range of scenarios (Fig. 1, ellipse 2). We represented low-frequency climatic effects on salmon productivity by varying the  $a$  parameter in the Ricker model (eq. 1) over time using one of four different hypothetical patterns, each one producing a different type of climate-induced scenario of variability in productivity (Fig. 2). (1) CONST: the initial  $a$  parameter value in eq. 1 was drawn from a normal distribution with a mean of 1.75 and standard deviation,  $\sigma_a$ , of 0.5 but was then held constant over the entire simulation. (2) STEP:  $a$  followed a step function with a period of 60 years, representing two alternating climatic re-

**Fig. 2.** Example time series of the Ricker  $a_t$  (productivity) parameter (eq. 1) for the four types of simulated climate-driven productivity patterns: constant productivity (horizontal dashed line), step function (solid line), autoregressive-lag-1-year (AR(1)) process (dotted line), and step function plus AR(1) process (dot-dashed line). The two cases with AR(1) terms each show only one realization of the 500 Monte Carlo trials.



games of 30-year duration (based on Hare and Francis 1995). Although we kept the period fixed, the magnitude of change between successive regimes was allowed to vary, which produces time series patterns that more closely resemble historical patterns than a fixed step-size pattern. At each step-down point, a new  $a$  value was drawn from a normal distribution with mean of 1.5 and standard deviation of 0.5. At each step-up point, a new  $a$  value was drawn from a normal distribution with a mean of 2.0 and a standard deviation of 0.5, resulting in an average step size of 0.5. In the 30 years between steps,  $a$  values were held constant (Fig. 2). (3) AR1:  $a$  was modeled by an autoregressive-lag-1-year AR(1) process. The initial  $a$  value,  $a_{\text{base}}$ , was calculated as in the CONST model above. Annual  $a_t$  values were calculated as

$$(2) \quad a_t = a_{\text{base}} + \Delta a_{t-1} \text{ where } \Delta a_t = \varphi \Delta a_{t-1} + z_t$$

The autocorrelation coefficient  $\varphi$  was set to 0.8 and  $z_t$  was drawn from a normal distribution with a mean of zero and a standard deviation of 0.3 (i.e.,  $\sigma_a \sqrt{(1 - \varphi^2)}$ ) (Fig. 2). (4) STEP\_AR1:  $a$  followed a combination of step function and random walk. Calculations were the same as for AR(1), except we used the STEP underlying temporal pattern rather than the CONST case for  $a_{\text{base}}$  (Fig. 2). We believe that this STEP\_AR1 pattern most closely represents past variation in the northeastern Pacific Ocean.

The parameters for each of the above hypothesized productivity patterns were selected to produce qualitatively realistic values for salmon productivity that emulate the natural patterns observed in our spawner–recruit data. The autocorrelation in the AR(1) process for each pattern mimics the average autocorrelation observed in the time series of residuals that arises from fitting the Ricker model separately to the spawner–recruit data for each of the 37 sockeye stocks. To accommodate the wide range of fluctuations in  $a$  under the four climatic patterns, the overall average  $a$  value was set to 1.75, or approximately six recruits per spawner in the absence of density dependence. This is somewhat higher than the overall data average of  $a = 1.5$  but still well within the range of long-term average values derived for individual sockeye salmon stocks (see Dorner et al. 2008).

The climate and environment module also included spatial structure to permit comparisons of hierarchical stock assessment models that can use such information with those that cannot. Spatial coherence (i.e., spatial autocorrelation or co-variation across space) in climatically driven patterns in stock-specific  $a$  values across the simulated area was created by replacing the normal distributions used in the four climate scenario generators (cases 1–4 above) with multivariate normal distributions and drawing deviates concurrently for all stocks. The multivariate normal distributions had covariance

$$(3) \quad \text{Cov}(i, j) = V e^{-\rho d(i, j)}$$

where  $d(i, j)$  is the distance between ocean-entry points of stocks  $i$  and  $j$ ,  $V$  is the overall variance in the parameter for which a sample is to be drawn, and  $\rho$  affects the covariance at each distance between stocks; the larger the value of  $\rho$ , the greater the reduction in spatial similarity with increasing distance between stocks. Thus, for example, for a CONST productivity pattern, the vector of initial  $a$  values (one for each of the 15 simulated stocks) was drawn from the multivariate normal distribution, with  $V$  describing the desired variance in Ricker  $a$  values among stocks. Nearby stocks would typically have more similar  $a$  values than stocks farther apart, with the degree of spatial similarity across a given distance determined by  $\rho$ .

In our sensitivity analyses, we used eq. 3 to model three levels of spatial coherence among stock productivities: NONE, CONFINED, and BROAD. NONE was the null model in which each stock's time series of Ricker  $a_t$  values was generated independently by the model. The CONFINED and BROAD cases were derived from data on sockeye salmon, specifically the spatial pattern of residuals from stock-specific fits of the Ricker spawner–recruit model (as estimated empirically by Peterman et al. 1998). The CONFINED scenario ( $\rho = 1.5$ ) was parameterized using data from Chignik and Bristol Bay sockeye salmon stocks and the BROAD scenario ( $\rho = 0.2$ ) reflected southern British Columbia sockeye stocks.

We also simulated environmental covariates. For the purpose of this paper, we define the term “environmental cova-

riate” to mean an indicator that is driven at least in part by the same (but not directly observable) process that drives spatio-temporal patterns of salmon productivity. Thus, in our case, the simulated environmental covariate contains useful information about the hidden drivers of salmon productivity, the effects of which are represented by the four patterns in productivity time series,  $a_t$ . To evaluate the performance of salmon forecasting and stock assessment models that use environmental covariates as independent variables (e.g., models that include sea surface temperature such as in Mueter et al. 2002a), we generated a simulated environmental covariate time series  $X_t$  in each Monte Carlo trial. This series of  $X_t$  values was designed to be correlated with the time series of  $a_t$  values input into the salmon population model:

$$(4) \quad X_t = \lambda a_t + e_t$$

where  $\lambda$  is a scaling factor,  $e_t \sim N(0, \sigma_e^2)$ , and

$$(5) \quad \sigma_e = (\lambda \sigma_a \sqrt{1 - r^2})/r$$

where  $\sigma_a$  is the standard deviation of the  $a_t$  time series for which the environmental covariate is to be created and  $r$  gives the desired degree of correlation between  $a_t$  and  $X_t$ . As with the  $a_t$  time series, we allowed for spatial coherence in  $X_t$ , which was again introduced by replacing the normal distribution for the error term  $e_t$  in eq. 4 with a multivariate normal distribution.

Parameter values used to simulate the environmental covariate were chosen to match the characteristics of time series of regional sea surface temperature anomalies (Mueter et al. 2002b). We chose sea surface temperature anomalies as the real-world example for our environmental covariate because previous analyses showed that including these anomalies improved the performance of forecasting models for a wide range of sockeye stocks (Mueter et al. 2002a, 2005; Su et al. 2004). To maintain consistency with the spatial coherence scenarios above, we used sea surface temperature anomalies from the Chignik and Bristol Bay areas and southern British Columbia, respectively, to parameterize our scenarios. Specifically, the  $\rho$  parameter in eq. 3 describing the extent of spatial coherence in the environmental covariate was set to 0.3 for the CONFINED scenario and was derived from the Chignik and Bristol Bay data, whereas  $\rho$  for the BROAD scenario was set to 0.6 based on southern British Columbia data. The  $r$  parameter in eq. 5 was selected to emulate the observed correlation between reconstructed  $a_t$  time series using a Kalman filter (Dorner et al. 2008) and sea surface temperature in the two areas:  $r = 0.6$  for the Chignik and Bristol Bay area and  $r = 0.3$  for southern British Columbia. However, rather than tying strength of correlation between  $a_t$  time series and the environmental covariate in with the CONFINED and BROAD scenarios of spatial coherence, we allowed the  $r$  parameter to vary independently of spatial coherence, thus creating a STRONG ( $r = 0.6$ ) and a MODERATE ( $r = 0.3$ ) scenario of correlation between  $a_t$  and the environmental covariate. To reemphasize, the environmental covariate in our model did not drive the temporal and spatial variation in the productivity parameters,  $a_t$ , of salmon stocks; those  $a_t$  values were generated separately. The environmental covariate was merely an in-

dex of such changes that could be used by some stock assessment models.

### Data collection

The data collection module (Fig. 1, ellipse 3) emulated the process of data collection by assembling a time series of “observed” quantities, including numbers of recruits returning annually, escapements to spawning grounds, and environmental covariates. Because there were no suitable data available at the scale of our model for quantifying errors introduced into spawner–recruit data by sampling and data collection protocols, we did not attempt to represent the effect of sampling errors explicitly. Instead, such errors were captured through the  $v_t$  term in the population model (eq. 1), which expresses deviation of recruits from the hypothesized biological relationship due to various factors not directly represented in the model (both natural variation and observation error).

Since fisheries biologists and managers most typically work with spawner–recruit time series that span a relatively short time period, the data collection module only retained the most recent 30 years of simulated data, thereby restricting the data record available in each simulation year to subsequent CLIM2 modules to a realistic temporal duration.

### Stock assessment and forecasting models

We implemented seven different types of stock assessment models (Table 2) and annually estimated their parameters for each of the 15 simulated salmon stocks (Fig. 1, ellipses 4 and 5). All stock assessment models had at their core the Ricker model (eq. 1) to describe the basic relationship between recruits and spawners, and each model was fit to the simulated spawner–recruit time series every year to obtain estimates of the  $a$  and  $b$  parameters, which were used in the management decision module described below to set annual escapement goals. While some stock assessment models were fit to data on each stock separately, others were able to integrate information from multiple stocks and (or) environmental covariates (Table 2 summarizes models and fitting methods). The single-stock models, fit independently to each stock, included BASIC, the single-stock Ricker model as in eq. 1, AR1, the BASIC Ricker model modified to allow for temporally autocorrelated variation (i.e., the error term was modeled as an AR(1) series as in eq. 2), and KF (Kalman filter), which was the BASIC Ricker model with time-varying  $a$  parameter fit using a Kalman filter procedure with a random-walk system equation (Peterman et al. 2000).

The multistock models, which were fit to all stocks simultaneously, included several Hierarchical Bayesian Models (HBMs) (Table 2; Su et al. 2004). In each case, the objective was to capitalize on similarities in productivity patterns across stocks,  $i$ , as expressed in the Ricker  $a_i$  parameters. The Ricker  $b_i$  parameters were treated as “fixed” (Gelman et al. 2004) and were assumed to vary independently for each stock. All HBMs were fit assuming autocorrelated residual errors described by an AR(1) process and a fixed effect for  $\varphi_i$ , which determines the strength of the temporal correlation in residuals. The HBMs included in the CLIM2 model (Table 2) were as follows. (1) HBM, a basic HBM; the Ricker  $a_i$  parameters were modeled as a random effect,

**Table 2.** The seven stock assessment models used in CLIM2.

Model	Description	Equation	Error term $\varepsilon_t$	Parameters fitted	Fitting method	Notes
BASIC	Single-stock Ricker model	$\log_e(R_t/S_t) = a - bS_t + \varepsilon_t$	$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$	$a, b$	Linear regression	
AR1	Single-stock Ricker model with autocorrelated errors	$\log_e(R_t/S_t) = a - bS_t + \varepsilon_t$	$\varepsilon_t = \varphi\varepsilon_{t-1} + e$ $e \sim N(0, \sigma_e^2)$	$a, b, \varphi$	Linear regression with grid search for $\varphi$	
KF	Single-stock Ricker model with time-varying $a$ parameter	$\log_e(R_t/S_t) = a_t - bS_t + \varepsilon_t$ $a_t = a_{t-1} + w_t$	$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ $w_t \sim N(0, \sigma_w^2)$	$a_t, b, \sigma_\varepsilon, \sigma_w$	Kalman filter assuming random walk for $a_t$	See Peterman et al. (2000) for details
HBM	Multistock hierarchical Bayesian Ricker model	$\log_e(R_{it}/S_{it}) = a_i - b_iS_{it} + \varepsilon_{it}$	$\varepsilon_t = \varphi\varepsilon_{t-1} + e$ $e \sim N(0, \sigma_e^2)$	$a_i$ : random effect $b_i$ : fixed $\sigma_e^*$ $\varphi$ : fixed	MCMC algorithm	See Su et al. (2004) for details
SHBM	Multistock hierarchical Bayesian Ricker model with spatial covariation	$\log_e(R_{it}/S_{it}) = a_i - b_iS_{it} + \varepsilon_{it}$	$\varepsilon_t = \varphi\varepsilon_{t-1} + e$ $e \sim N(0, \sigma_e^2)$	$a_i$ : CAR $b_i$ : fixed $\varphi$ : fixed $\sigma_e^*$	MCMC algorithm	See Su et al. (2004) for details
HBM-ENVIRO	Multistock hierarchical Bayesian Ricker model with environmental covariate	$\log_e(R_{it}/S_{it}) = a_i - b_iS_{it} + c_iX_{it} + \varepsilon_{it}$	$\varepsilon_t = \varphi\varepsilon_{t-1} + e$ $e \sim N(0, \sigma_e^2)$	$a_i$ : random effect $b_i$ : fixed $c_i$ : random effect $\varphi$ : fixed $\sigma_e^*$	MCMC algorithm	See Su et al. (2004) for details
HBM-SENVIRO	Multistock hierarchical Bayesian Ricker model with environmental covariate and spatial covariation	$\log_e(R_{it}/S_{it}) = a_i - b_iS_{it} + c_iX_{it} + \varepsilon_{it}$	$\varepsilon_t = \varphi\varepsilon_{t-1} + e$ $e \sim N(0, \sigma_e^2)$	$a_i$ : CAR $b_i$ : fixed $c_i$ : CAR $\varphi$ : fixed $\sigma_e^*$	MCMC algorithm	See Su et al. (2004) for details

**Note:** Subscript  $t$  refers to brood year and  $i$  refers to the stock,  $i = 1$  to 37.  $R_t$ , observed total recruits from brood year  $t$  (i.e., from all age groups combined);  $S_t$ , observed spawning escapement in year  $t$ ;  $R_{it}$ , observed total recruits from brood year  $t$  for stock  $i$ ;  $S_{it}$ , observed spawning escapement in year  $t$  for stock  $i$ ;  $X_t$ , observed environmental covariate relevant to brood year  $t$ ; CAR refers to a Gaussian conditional autoregressive distribution.

\*Since HBM parameters are described as probability distributions, the Markov Chain Monte Carlo (MCMC) algorithm actually fits the parameters that describe the specified distribution, e.g., for a random effect, this would include the parameters of the common hyperdistribution.



i.e.,  $a_i$  values for the various stocks were assumed to originate from a single normal distribution (Gelman et al. 2004). (2) SHBM, a basic HBM with an added provision for spatial covariation in the  $a_i$  values by giving the  $a_i$  a spatially correlated prior distribution based on the Gaussian conditional autoregressive (CAR) structure (Besag et al. 1991; Banerjee et al. 2004; Su et al. 2004). (3) HBM-ENVIRO, a basic HBM with an environmental covariate; both the  $a_i$  parameters and the  $c_i$  parameters associated with the environmental covariate were modeled as normally distributed random effects. (4) HBM-SENVIRO, a version of HBM-ENVIRO that explicitly incorporates spatial covariation; both the  $a_i$  parameters and the  $c_i$  parameters associated with the environmental covariate were modeled assuming CAR distributions, i.e., similarities in parameter values were assumed to decrease with increasing distance between stocks.

All HBMs were fit via a Markov Chain Monte Carlo (MCMC) algorithm to establish the shape of the respective posterior distributions (and, where applicable, hyperdistributions) for the key parameters  $a_i$ ,  $b_i$ , and  $c_i$  as well as parameters describing the strength of spatial covariation, temporal autocorrelation in residuals, and residual variance. For further information on the priors used for each parameter and details on model formulation and the MCMC algorithm, see Su et al. (2004).

As we discuss later, one cause of OU (deviations between target and realized spawning escapements) is the error in forecasts of abundance of adult recruits. We therefore used a total of 10 models to forecast returns for a given year from recruitment and spawner data from previous years. All seven of the above stock assessment models doubled as forecasting models. Once their parameters were estimated, we used them to calculate an estimate of returns from observed spawner abundance from previous years. In addition to these seven stock assessment models, we evaluated three forecasting models described in the next paragraph that predicted returns without fitting any type of Ricker spawner–recruit model and thus could not be used in stock assessments to estimate the Ricker  $a$  and  $b$  parameters. However, because the management module requires these parameter estimates, we used the BASIC Ricker model as the stock assessment model to produce those estimates when we evaluated the three forecasting models that do not use the Ricker  $a$  and  $b$  parameters.

The three forecasting models that predicted returns without fitting a spawner–recruit relationship included the following. (1) LAG4, a model that simply estimates returns for the current year equal to returns 4 years previously, i.e.,  $R_{yr} = R_{yr-4}$ , where  $R_{yr}$  is the number of recruits forecasted for the current year and  $R_{yr-4}$  is the number of recruits that returned 4 years earlier. We selected the lag-4 model over similar nonparametric models that forecast abundance of recruits from previous recruits because, based on retrospective analysis of observed returns from the 37 sockeye stocks, it performed better than other lagged or weighted-average models (Haeseker et al. 2008). (2) SIB, a model that estimates abundance of recruits of some age stanzas from recruits of siblings of the same brood year that return 1 year earlier:

$$(6) \quad \log_e(R_{x,yr}) = c + d \log_e(R_{x-1,yr-1}) + \varepsilon_{yr}$$

where  $R_{x,yr}$  is the number of later-returning siblings aged  $x$  years returning in year “yr”,  $R_{x-1,yr-1}$  is the number of siblings spending 1 year less in the ocean and hence returning 1 year earlier,  $c$  and  $d$  are parameters describing the historically observed relationship between the two sibling age groups, and  $\varepsilon_{yr} \sim N(0, \sigma_\varepsilon^2)$  (see Peterman 1982). In the CLIM2 simulations, the sibling model was fit to the time series of simulated returns and was used for predicting abundance of those age classes for which earlier-returning siblings were available. Returns for the first age stanzas were predicted using a model identical to the LAG4 model above except applied only to returns for the specific age stanza. (3) HYBRID, a “hybrid” sibling model that uses either the LAG4 forecast (the default) or the forecast from the SIB model, contingent on the magnitude of the variance in residuals from the sibling relationship fit to the simulated data (see Haeseker et al. 2007). This model only uses the SIB model’s forecast when it fits the data well, as indicated by a residual variance in the sibling relationship that is below a given variance cutoff. A pilot analysis with simulated returns, following the method described by Haeseker et al. (2007), suggested an optimal variance cutoff of 2.0 for the simulated data, which was well in line with the range of 0.7–3.0 derived by Haeseker et al. (2007) for the observed returns from the 37 sockeye stocks.

With the exception of the LAG4 model, all of the 10 forecasting models predicted the logarithm of recruit abundance rather than actual recruit abundance. These forecasts were thus transformed from log scale back to recruit abundances by correcting for the well-known bias associated with transforming lognormally distributed variables (Hayes et al. 1995):

$$(7) \quad R_{pred} = \exp\left(R_{pred}^* + \frac{\sigma^2}{2}\right)$$

where  $R_{pred}^*$  is the predicted recruit abundance in log scale,  $\sigma^2$  is the variance of residuals estimated from fitting the particular forecasting model in question, and  $R_{pred}$  is the resulting prediction of recruit abundance on an arithmetic scale. Depending on the model,  $R_{pred}^*$  and  $R_{pred}$  may refer either to predictions for an individual age stanza (for the sibling models) or to predictions of total recruit abundance (for the rest of the stock assessment and forecasting models).

In addition to the stock assessment and forecasting models listed above, scenarios were also run with PERFECT information to investigate the potential benefits of completely eliminating errors and bias in forecasts and parameter estimates. Such PERFECT scenarios used actual population parameter values and actual recruitments instead of model estimates.

### Management decision making

We simulated decision making (Fig. 1, ellipse 6) by calculating the annual target escapement from the most recent annual estimates of Ricker  $a$  and  $b$  parameters and the residual variance of the fitted stock assessment model by using a previously derived four-parameter S-shaped relation (Peterman et al. 2000):

$$(8) \quad S_{\text{target}} = \frac{a'}{b} \left( K_1 - \frac{K_2 a'^{K_4}}{K_3 + a'^{K_4}} \right) \text{ with } a' = a + \frac{\sigma^2}{2}$$

where  $S_{\text{target}}$  is the target escapement,  $a$  and  $b$  are the estimates of Ricker  $a$  and  $b$  parameters,  $a'$  is an adjustment of  $a$  to account for lognormal variation (Hilborn 1985),  $\sigma^2$  is the residual variance of the model fit, and shape parameters are set to  $K_1 = 0.5$ ,  $K_2 = 0.65$ ,  $K_3 = 8.7$ , and  $K_4 = 1.27$  (Peterman et al. 2000). Target escapements were calculated independently for each stock every year using the parameters supplied by the stock assessment model being evaluated at the time. In a subset of our sensitivity analyses, we also tested the effect of applying a management safety margin to reflect a more biologically conservative action. To do this, the model increased the target escapement that was initially calculated from eq. 8 by a safety margin of either 20% or 40%. Incidentally, we did not evaluate Walters and Parma's (1996) strategy of fixed-percent harvest rates here because in the simpler, single-stock CLIM model, Peterman et al. (2000) found that fixed harvest rates generated the lowest average catch of all management strategies considered; the same would be true here for the 15-stock case.

### Harvesting, including OU

The harvesting module (Fig. 1, ellipse 7) includes OU, which describes deviations between the actual realized escapement and the target escapement goal. The term "outcome uncertainty" extends the more often used narrower concept of "implementation error" by recognizing that deviations between management targets and realized outcomes are the net result of changes in catchability coefficients related to physical and biological dynamics as well as noncompliance with fishing regulations (Holt and Peterman 2006). We simulated this stochastic OU by calculating  $\Delta S$ , the deviation between the realized and target escapements each year:

$$(9) \quad \Delta S = \log_e(S_{\text{actual}}) - \log_e(S_{\text{target}}) \\ = f + g \log_e(R_{\text{actual}}) + h[\log_e(R_{\text{actual}}) - \log_e(R_{\text{pred}})] + \varepsilon$$

where  $S_{\text{actual}}$  and  $S_{\text{target}}$  are the actual and target escapements, respectively, and  $S_{\text{target}}$  was either as estimated by eq. 8 or the new  $S_{\text{target}}$  after its modification by a safety margin.  $R_{\text{actual}}$  and  $R_{\text{pred}}$  are actual and predicted returns,  $f$ ,  $g$ , and  $h$  are parameters estimated from historical time series of predicted and actual returns and resulting escapements as described below, and  $\varepsilon \sim N(0, \sigma_\varepsilon^2)$  describes the variability in OU. Catch in CLIM2 was then calculated as the difference between the actual recruitment simulated in module 1 of the operating model and the actual escapement estimated using a rearranged version of eq. 9.

Empirical evidence for sockeye salmon (Holt and Peterman 2006, 2008) shows that, depending on the stock, there are two generic patterns of OU, both of which can be described using eq. 9. TYPE-1 OU depends only on forecast error, whereas TYPE-2 OU depends on that error as well as on total abundance of adult recruits,  $R_{\text{actual}}$ . To determine the potential impact of reducing or eliminating OU, we also ran sensitivity analyses without any OU.

Parameters for TYPE-1 OU were based on data from the Fraser River, British Columbia, sockeye fishery. For the

Early Stuart, Early Summer, and Summer runs of Fraser sockeye, OU has historically been primarily influenced by preseason prediction error (Holt and Peterman 2006). By fitting eq. 9 to the combined data for these three run timing groups, we derived parameter estimates for TYPE-1 OU as  $f = 0.5$ ,  $g = 0$ ,  $h = 0.4$ , and  $\sigma_\varepsilon = 0.4$ .

Parameters for TYPE-2 OU were based on data from the Bristol Bay, Alaska, sockeye fishery. Bristol Bay sockeye return over a shorter period (by a few weeks) than Fraser River sockeye. Although there is some limited opportunity to use revised "in-season" estimates of run size to adjust management targets in Bristol Bay, the short fishing season and remote location mean that there is limited potential for adjustments in commercial harvesting capacity in response to changes in management targets. Escapements in the Bristol Bay area historically tended to be below the target when returns were low and were higher than the target when returns were high. The parameters for TYPE-2 OU thus include a run-size effect, i.e.,  $f = 0.1$ ,  $g = 0.3$ ,  $h = 0.4$ , and  $\sigma_\varepsilon = 0.6$ .

The net effect of these dynamic model components on OU is twofold. First, the run-size effect creates a tendency towards underescapement (abundance of spawners lower than the target escapement) at low abundances and overescapement at high abundances for TYPE-2 OU. Second, TYPE-1 OU produces more frequent cases of overescapements than TYPE-2 OU over the range of recruitments simulated by our model (example in Fig. 3). This figure shows results for two 100-year Monte Carlo trials, both with the same series of random numbers. Although these are only a small sample of the hundreds of simulated trials, they show that, in this case, TYPE-1 OU led to overescapement in 83 of the 100 years, whereas TYPE-2 OU did so in only 57 years.

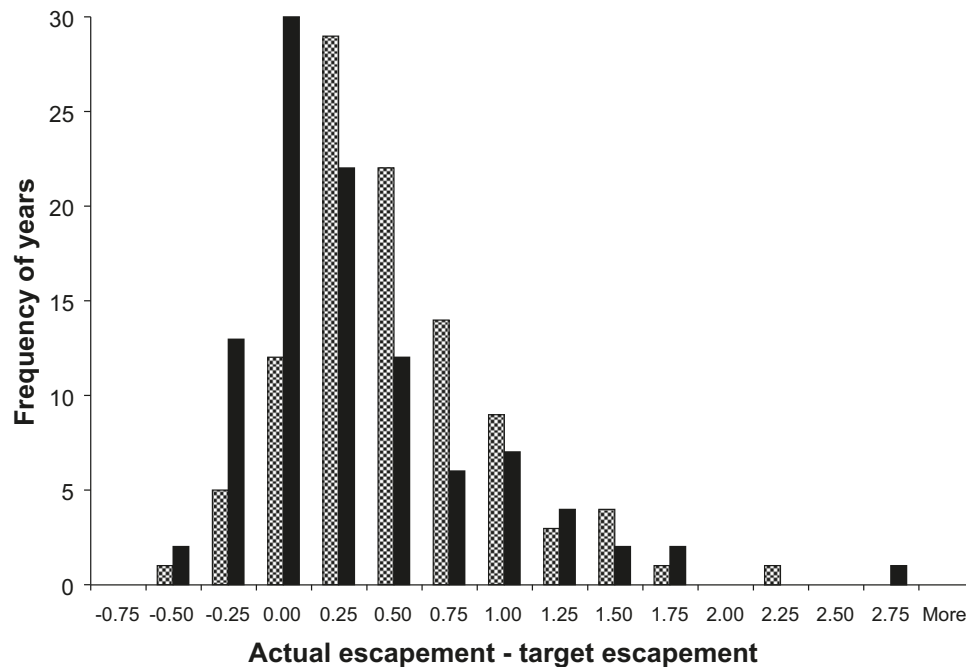
### Model evaluation

Previous sections described the numerous assumptions that we explored about the natural and management systems. In total, we simulated 1516 different scenarios composed of three levels of correlation between the environmental covariate and stock productivity, three magnitudes of spatial coherence among stocks in their productivity, three safety margins on escapement goals, three types of OU, four temporal patterns of climate-induced changes in productivity, seven different stock assessment models, and 10 forecasting models.

CLIM2 was run with 15 simulated salmon stocks, which were randomly distributed along a fictional coastline of unit length by drawing 15 random along-shore locations between 0 and 1 from a uniform distribution. The number of stocks and their locations were held constant across all scenarios and model runs. Trial runs indicated that variance in key model performance measures stabilized after approximately 400–500 Monte Carlo replicates, so we ran each scenario 500 times.

Each individual CLIM2 Monte Carlo replicate started by generating an initial population of spawners (see Salmon populations section above) followed by a 20-year initialization period to create enough data to fit stock assessment and forecasting models (14 pairs of spawner and recruit data at this point because for the remaining six data years, complete abundance estimates for recruits were not yet available due

**Fig. 3.** Frequency distribution for actual escapement abundance minus target escapement for TYPE-1 (cross-hatched bars) and TYPE-2 (solid bars) outcome uncertainties. There were 100 simulated years of Monte Carlo trials for each type of outcome uncertainty. Runs without any outcome uncertainty would all fall in the bin for zero difference between actual and target escapement.



to the age structure of returns). During this initialization period, the dynamics of the spawning population were simulated fully, but the management components of CLIM2 were disabled and a fixed annual harvest rate of 0.5 was applied instead to all stocks. The initialization period was followed by 100 simulated years during which the full management system was enabled and statistics on various aspects of model performance were collected.

For each run, CLIM2 recorded the average, variance, and coefficient of variation for annual harvest and annual escapement plus an index of conservation concern that tallied the proportion of years in which spawner abundance fell below 10% of the unfished equilibrium abundance (i.e., “conservation-concern index”). Additional performance measures included root mean square error (RMSE) and bias in the  $a$  and  $b$  parameter estimates produced by a stock assessment model. We also calculated forecasting error (RMSE of forecast compared to actual run size) and forecasting bias in run-size estimates produced by the forecasting model (average difference of forecasted abundance minus actual run size). The latter two performance indicators reflect the ability of the chosen forecasting model to estimate adult returns.

## Results

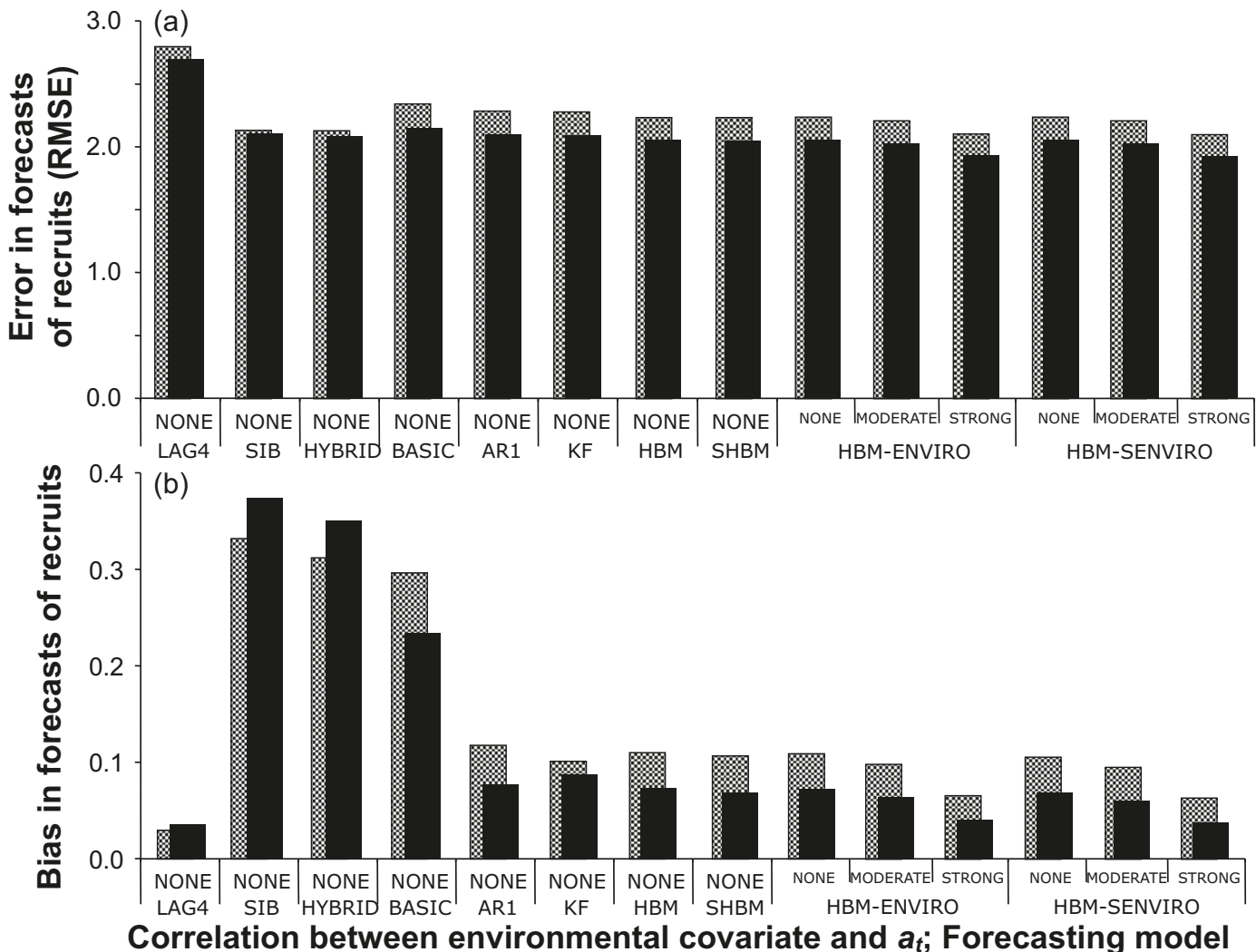
We found that the type of climate-driven productivity pattern, type of OU, and magnitude of safety margin each affected performance of the stock assessment and forecasting models as well as average catch and the conservation-concern index. We first present results for the “base case” consisting of a STEP\_AR1 productivity pattern, CONFINED (i.e., typical) spatial coherence among stocks in their Ricker  $a$  parameters, and a safety margin of zero because these are

the settings that we believe are most applicable to the majority of populations and to historical management for salmon in the eastern Pacific. Unless stated otherwise, results are shown for all forecasting/stock assessment models, both types of OU, and the three levels of correlation between environmental variables and Ricker  $a$  parameters. Following the base case results, we report sensitivity analyses of effects of different assumptions about productivity patterns, spatial coherence, and safety margins.

### Base case: errors and bias in estimates of forecasting and stock assessment models

We first present performance indicators of the 10 models used to forecast abundance of adult salmon recruits. Below, abundance figures should be interpreted with reference to the model’s equilibrium abundance of  $R = 1.75$ , determined as the ratio of the  $a$  and  $b$  parameters of the Ricker model, or 1.75/1 in our case. All models outperformed the “naïve” LAG4 model by producing smaller forecasting error (Fig. 4a), but the LAG4 model was among the models with the smallest forecasting bias (Fig. 4b). Except for the LAG4 model, differences among models in forecasting error were relatively small, with <11% difference between the best and worst models for TYPE-1 OU and <5% for TYPE-2 OU (Fig. 4a). However, differences among models in forecasting bias were more substantial. Compared with the worst model (SIB), the best model (LAG4) showed a 90% reduction in bias (Fig. 4b). Among single-stock Ricker-type models, the addition of the AR1 error term (AR1 model) and the time-varying  $a_t$  parameter (KF model) both helped reduce forecasting error (by 2%–3%) as well as forecasting bias (by approximately 60%–70%) compared with the BASIC Ricker model. Formulating the Ricker model in a multistock context also helped improve model performance. Specifically,

**Fig. 4.** Root mean square error (RMSE) in (a) forecasts of abundance of adult recruits and (b) bias in those forecasts (forecast abundance minus actual run size averaged over all simulated years) for 10 forecasting models (bottom labels on x-axes) and two types of outcome uncertainty: TYPE-1 OU (cross-hatched bars) and TYPE-2 OU (solid bars). Results for HBM-ENVIRO and HBM-SENVIRO, the two models that include an environmental covariate, are shown for three levels of “true” strength of correlation between that covariate and the Ricker  $a_t$  parameter (top labels on x-axes). All results are for the base case scenario defined by a STEP\_AR1 productivity pattern, CONFINED spatial coherence for productivity (Ricker  $a_t$ ), and a safety margin of zero.



HBM models generally had 4%–10% lower forecasting error than the Ricker-type single-stock models (Fig. 4a), and forecasting bias was similar to, or lower than, that for the AR1 and KF models and approximately 60%–80% lower than that for the BASIC Ricker model (Fig. 4b). Forecasting error for the two sibling models was slightly lower than for the BASIC Ricker model and was comparable with that for the AR1, KF, and HBM models (Fig. 4a), although the sibling model forecasts ranked worst in terms of forecasting bias (Fig. 4b). The HYBRID sibling model was a slight improvement over the standard sibling model.

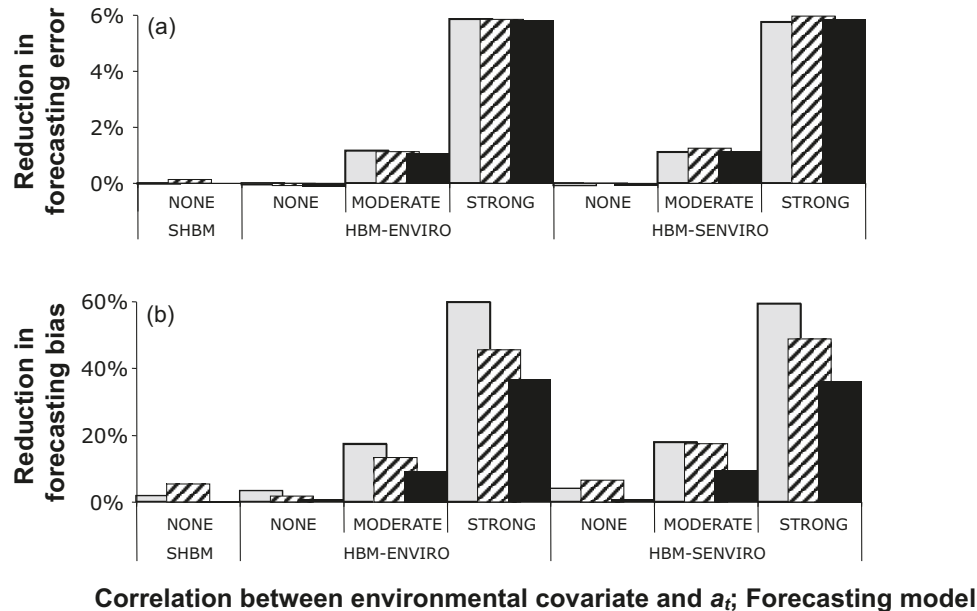
Adding an environmental covariate improved forecasting performance for HBM models, with the degree of improvement governed by the strength of correlation in the operating model between the environmental covariate and salmon productivity. Assuming MODERATE correlation, the HBM-ENVIRO and HBM-SENVIRO models, which included environmental covariates, both had approximately 1% lower

forecasting error than the corresponding simpler HBM model (Fig. 5a) and 9%–18% lower bias, depending on the spatial covariation of productivity across stocks (Fig. 5b), whereas forecasting error was reduced by approximately 6% and bias was reduced by 37%–60% in the STRONG environmental correlation scenarios (Fig. 5b). The magnitude of these benefits of the HBM-ENVIRO and HBM-SENVIRO models held for both types of OU.

As with addition of an environmental covariate, explicit allowance for spatial covariation in productivity across stocks in the SHBM and HBM-SENVIRO models helped to further improve forecasting performance, although the incremental effect was less pronounced than for the addition of an environmental covariate. The reduction in forecasting error when using the SHBM and HBM-SENVIRO models instead of the corresponding nonspatial HBM and HBM-ENVIRO models, respectively, was <1% (Fig. 5a), and reduction in forecasting bias was about 5% or less (Fig. 5b).



**Fig. 5.** Reduction (compared with the simpler corresponding HBM model) in (a) RMSE and (b) bias in forecasts of abundance of adult recruits when information on the spatial extent of high positive correlations among stocks and an environmental covariate were incorporated into the versions of the multistock hierarchical Bayesian model (bottom x-axis labels). Top x-axis labels refer to strength of correlation between the environmental covariate and the salmon productivity parameter ( $a_i$ ). Shadings of bars reflects the spatial extent of high positive correlations in low-frequency variability in productivity generated by eq. 3: shaded bars, NONE; hatched bars, CONFINED; solid bars, BROAD. Other components of scenarios were at their base case settings: STEP\_AR1 productivity pattern, a safety margin of zero, and TYPE-2 OU.



The most basic indicator of performance for stock assessment models is their ability to estimate the  $a$  and  $b$  parameters of the Ricker relationship. Multistock models generally produced more accurate parameter estimates than single-stock models, reducing RMSE in  $a$  by 7%–13% and in  $b$  by 6%–21% over the best single-stock model (Supplemental Table S1).<sup>4</sup> However, bias in the  $a$  parameter for multistock models was typically considerably larger than for the least-biased single-stock model (the BASIC Ricker model) (Fig. 6a). Also, whereas bias in  $b$  was negative for the BASIC Ricker model, bias in  $b$  was positive for all multistock models as well as for the AR1 and KF models and was the largest for the KF model (Fig. 6b). Although the LAG4, SIB, HYBRID, and BASIC scenarios all use the same BASIC Ricker model for stock assessment, bias in parameter estimates was different in each case. This is because the different forecasts influenced properties of the resulting spawner–recruit time series, thereby affecting the outcome of the fitting process. Note that in all cases in Fig. 6, however, biases in estimates of  $a$  and  $b$  were small in absolute terms (<5% difference from the mean true value for both  $a$  and  $b$ ).

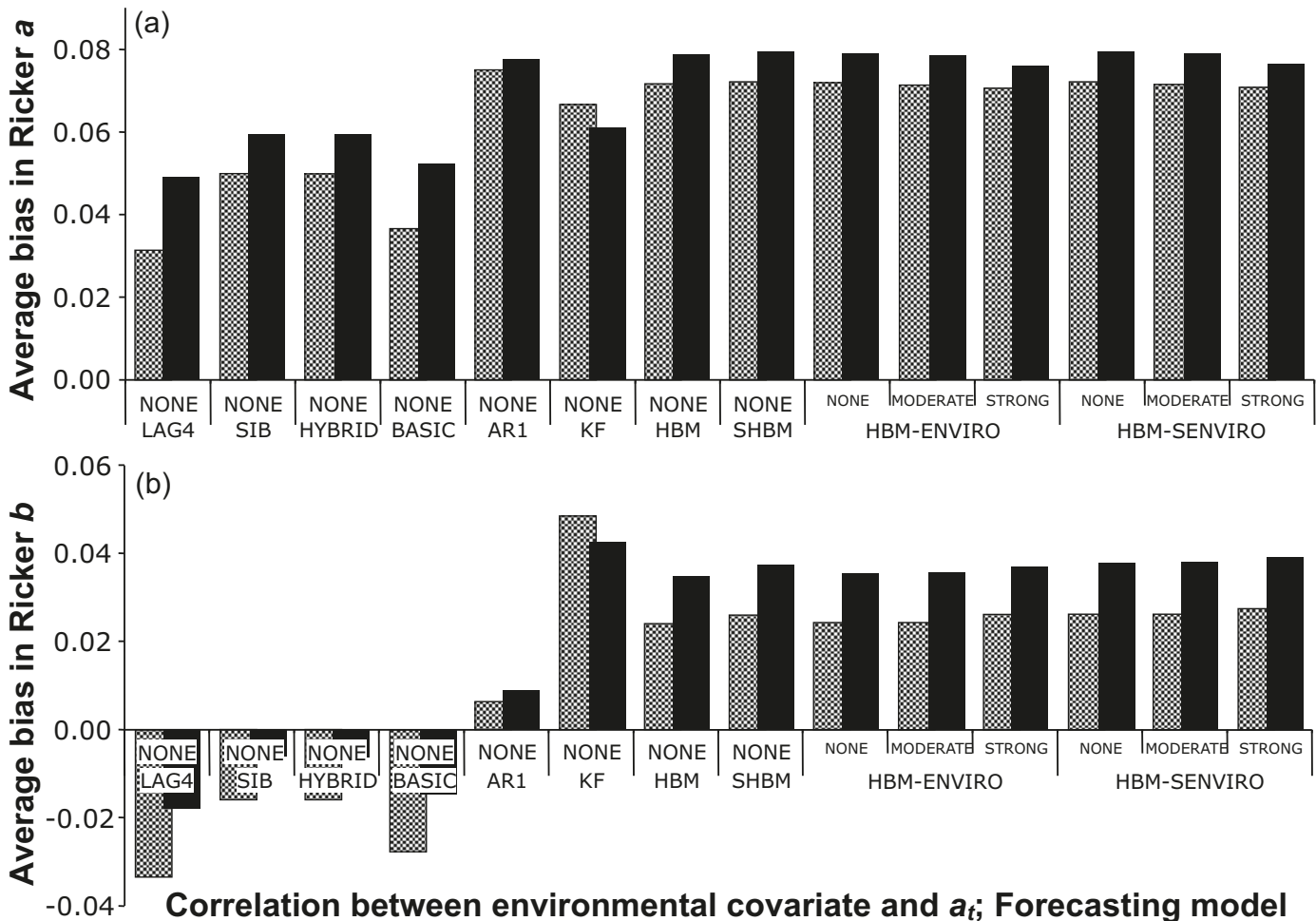
#### Base case: catch and conservation-concern index for forecasting and stock assessment models

While good forecasts of abundance of adult recruits can help increase average annual catch and lower the average index of conservation concern, forecasting performance by it-

self was not a good predictor of outcomes for catch and conservation-concern index (the latter was the proportion of years with low abundance, i.e., <10% of the unfished equilibrium). Instead, bias in parameter estimates and the resulting bias in escapement targets often had a stronger influence on outcomes than forecasting performance, and high values for average catch due to overly “aggressive” targets often went along with high values for the index of conservation concern and vice versa. For example, the LAG4 model produced poor forecasts (Fig. 4a) and comparatively low catches (Fig. 7a), but this model was well within the range of other models in terms of conservation-concern index for the TYPE-1 OU scenario, and it had the lowest index of conservation concern of all models for the TYPE-2 OU scenario (Fig. 7b). Conversely, the KF model produced comparatively good forecasts of adult recruits (Fig. 4) and had high overall catch but also had by far the highest conservation-concern index, likely due to the occasional large overestimates in the Ricker  $b$  parameter (Fig. 6b). Similarly, the multistock models, which performed well in terms of forecasting error and RMSE in parameter estimates, produced higher catches than the BASIC Ricker model (but only by about 1%–5%) but also had higher conservation-concern indices than the BASIC model for the TYPE-2 OU. These observed differences among the BASIC, KF, and multistock models result from the bias in parameter estimates, especially the  $b$  parameter (Fig. 6b). Positive bias in  $b$  biases es-

<sup>4</sup>Supplementary data for this article are available on the journal Web site (<http://cjfas.nrc.ca>) or may be purchased from the Depository of Unpublished Data, Document Delivery, CISTI, National Research Council Canada, Building M-55, 1200 Montreal Road, Ottawa, ON K1A 0R6, Canada. DUD 5292. For more information on obtaining material, refer to <http://cisti-icist.nrc-cnrc.gc.ca/eng/ibp/cisti/collection/unpublished-data.html>.

**Fig. 6.** Average bias in estimates of the (a) Ricker  $a$  parameter and (b) Ricker  $b$  parameter (defined as estimate minus true value) for two types of outcome uncertainty, TYPE-1 OU (cross-hatched bars) and TYPE-2 OU (solid bars), using the base case with STEP\_AR1 productivity, CONFined spatial coherence among stocks in their productivity (Ricker  $a$  parameters), and a safety margin of zero.



### Correlation between environmental covariate and $a_t$ ; Forecasting model

capement goals downward and catch too high. In contrast, use of either of the relatively good sibling models for forecasting along with the relatively low-bias BASIC Ricker model for stock assessment produced slightly higher catches (Fig. 7a) and at the same time lower conservation-concern index values than the BASIC model (Fig. 7b).

Except for the LAG4 model, differences among models in average catch were relatively small, i.e., within a 5%–6% range (Fig. 7a). Moreover, for all models except the LAG4 model, catch was within 85%–95% of the catch obtained for the PERFECT information scenario, which used actual population parameters and run size in place of model estimates (Fig. 7a).

#### Effect of OU

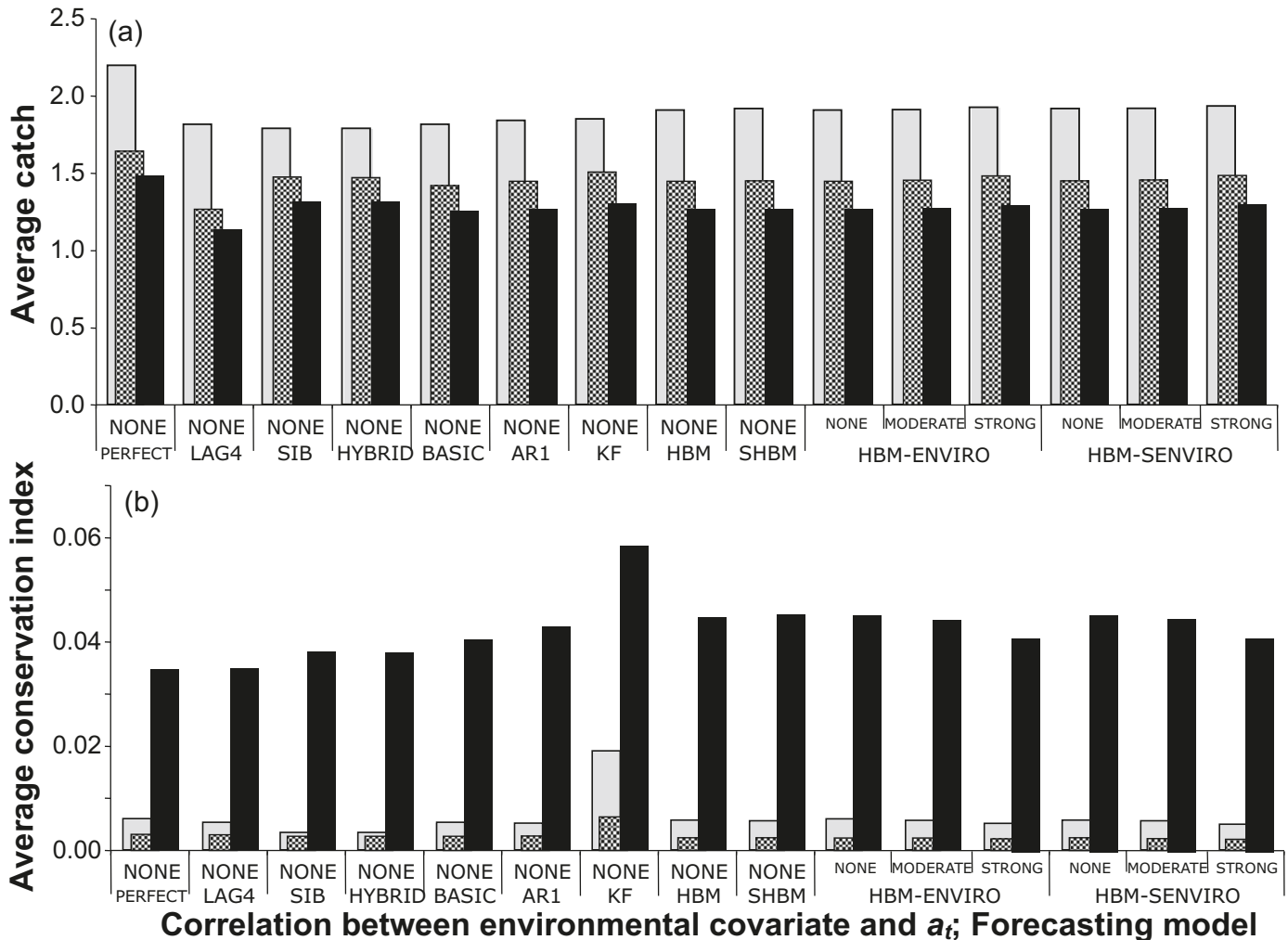
Removing OU altogether resulted in substantially higher catches (average of 23% above the case of TYPE-1 OU and 32% increase over TYPE-2 OU; Fig. 7a). These increases in average catch across stock assessment and forecasting models are quite close to the increases obtained for the PERFECT scenario, which produced 25% and 32% increases in catch above the TYPE-1 and TYPE-2 OU cases, respectively. These differences in average catch between cases with and without OU are attributable to the different varian-

ces in catches over time created by OU. Because the catch-reducing effect of OU was very similar across scenarios (including the PERFECT scenario that essentially used perfect forecasting and stock assessment models), improvements in forecasting or stock assessment models are unlikely to overcome this effect of OU to substantially increase catch if OU remains unchanged. However, reduction or modification of OU might result in substantial increases in long-term average catch.

Removing OU also generally resulted in a substantially lower index of conservation concern (proportion of stock years with escapements <10% of the unfished equilibrium) than for scenarios using TYPE-2 OU (Fig. 7b). Values of the conservation-concern index for TYPE-2 OU averaged seven times those of runs without OU. However, those values for TYPE-1 OU were on average about only half of what we observed for runs without OU, which we discuss later.

The effect of TYPE-2 OU (the one in which deviations between management target escapement and realized escapement were a function of two factors, total run size and the difference between forecast and actual abundance of recruits) was quite distinct from that of TYPE-1 OU (in which deviations were only affected by differences between the forecast and actual abundance of recruits). The forecasting

**Fig. 7.** (a) Average catch and (b) average conservation-concern index for the base case with STEP\_AR1 productivity, CONFINED spatial coherence among stocks in their productivity, and a safety margin of zero. Results are shown for three cases of outcome uncertainty: none (shaded bars), TYPE-1 OU (cross-hatched bars), and TYPE-2 OU (solid bars).



and stock assessment models generally performed better under the TYPE-2 OU than under the TYPE-1 OU; forecasting errors for run size (Fig. 4a) were smaller, as was forecasting bias (with the exception of the sibling models and the LAG4 model) (Fig. 4b). Nonetheless, as explained later, catch for TYPE-2 OU was consistently around 10% lower than for TYPE-1 OU (Fig. 7a), and the conservation-concern index for TYPE-2 OU was about 10–20 times that of TYPE-1 OU (Fig. 7b).

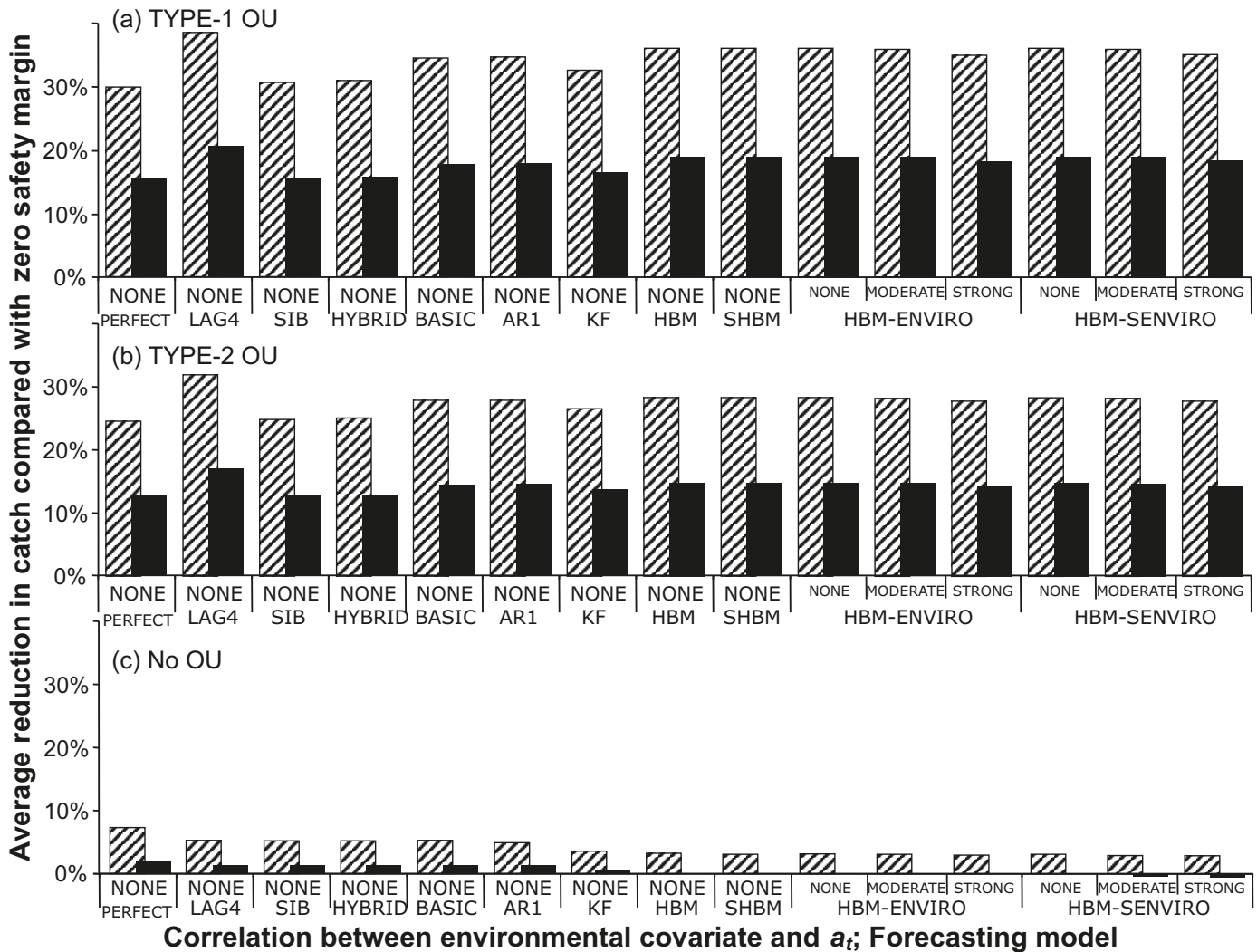
#### Effect of a safety margin

Not surprisingly, introduction of safety margins that increased target escapements consistently resulted in reduced catch. When we used TYPE-1 OU, the average reduction in catch across models (compared with cases without any safety margin) was 18% for a 20% safety margin and 35% for a 40% safety margin (Fig. 8a). For TYPE-2 OU, the corresponding average reductions were 14% and 28% (Fig. 8b). Most importantly from a management perspective, when we ran simulations without any OU, average catches were barely reduced (<4%) by using either of the safety margins on escapement (Fig. 8c).

Reductions in catch caused by nonzero safety margins were associated with concurrent large decreases in the index of conservation concern for TYPE-2 OU, i.e., by 37% and 56% on average across models for the 20% and 40% safety margins, respectively (Fig. 9b). For scenarios without OU, corresponding reductions were even higher, averaging 50% and 61% (Fig. 9c). Safety margins in the presence of TYPE-1 OU also resulted in a consistent, but much smaller, reduction of the conservation-concern index than with TYPE-2 OU (Fig. 9a).

When evaluating the effect of different safety margins on the conservation-concern index, it is important to keep in mind that values of the conservation-concern index for TYPE-1 OU are very low in the first place (<1% of stock years have escapements below 10% of the unfished equilibrium; Fig. 7b), even for scenarios without a safety margin. As well, the reductions achieved by introducing safety margins are quite small in absolute terms, even when relative percent gains over the scenario without safety margin are large. Safety margins are likely to be a more important management tool in a TYPE-2 OU setting, where values of the conservation-concern index exceed those for TYPE-1 OU by about sevenfold.

**Fig. 8.** Average percent reduction in catch (compared with the case with zero safety margin) resulting from two levels of safety margin on escapement goals, 40% (hatched bars) and 20% (solid bars), with (a) TYPE-1 OU, (b) TYPE-2 OU, and (c) no OU for the base case with STEP\_AR1 productivity and CONFINED spatial coherence among stocks in their productivity.



### Effect of climate-driven variability in productivity

#### Forecasting errors and bias

As expected, models performed best at forecasting abundance of recruits under constant productivity. Forecasting error was lowest (Fig. 10a) and forecasting bias was generally smallest (Fig. 10b) for the CONST pattern. Results shown in Fig. 10 and reported below are for TYPE-2 OU, but figures for TYPE-1 OU are generally very similar. Adding variability to  $a_t$  by applying a time-varying productivity regime increased the RMSE forecasting error over the CONST pattern by 14%–32% for the STEP pattern, 33%–45% for the AR1 pattern, and 42%–66% for the STEP\_AR1 pattern (Fig. 10a). Adding a STEP pattern to the productivity pattern resulted in a positive forecasting bias in all cases (Fig. 10b). Adding an AR(1) pattern resulted in a slightly negative forecasting bias for the models that fit the residuals of the Ricker stock–recruit relationship as an AR(1) process (AR1 and the various HBM models), whereas for the other models, an AR1 pattern resulted in positive forecasting bias.

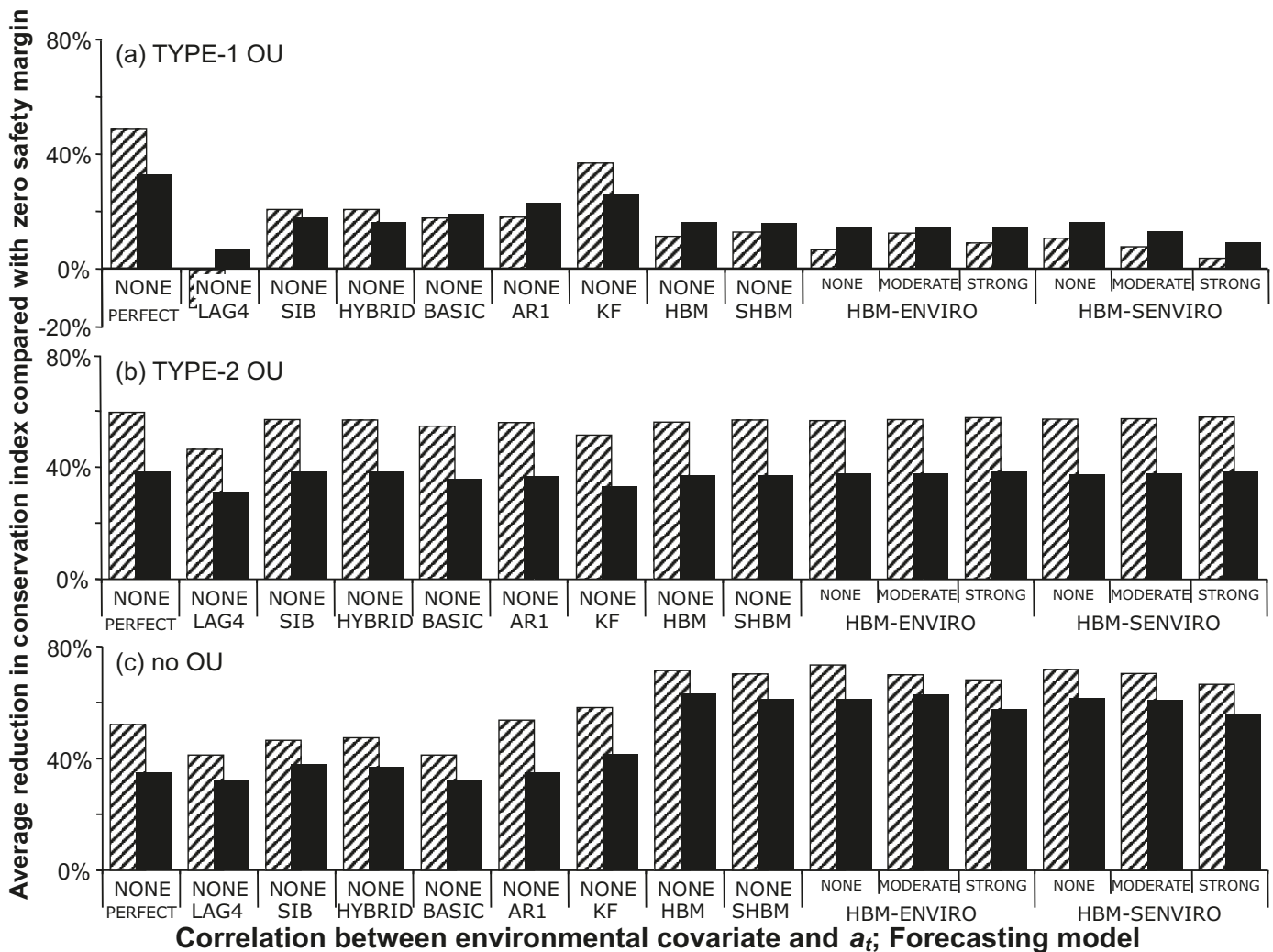
The STEP\_AR1 pattern, which combined the STEP and AR1 patterns (Fig. 2), resulted in reduced forecasting bias compared with that of the STEP pattern for models that fit Ricker residuals as an AR(1) process and increased forecasting bias over the STEP pattern for the other models (Fig. 10b).

#### Catch and conservation-concern index

These increases in forecasting error in response to introducing climate-driven variability did not necessarily result in less desirable outcomes for catch and conservation-concern index compared with the constant-productivity case. Instead, introducing a STEP pattern resulted in only a slight reduction in catch but a larger reduction in conservation concern compared with the CONST scenario (Fig. 11), whereas addition of an AR1 pattern increased both conservation-concern index and catch for TYPE-2 OU (Fig. 11). Results were similar for TYPE-1 OU. The STEP\_AR1 pattern produced an effect intermediate between that of the STEP and AR1 patterns (Fig. 11).



**Fig. 9.** Average percent reduction in conservation-concern index (compared with the case with zero safety margin) resulting from two levels of safety margin on escapement goals, 40% (hatched bars) and 20% (solid bars), with (a) TYPE-1 OU, (b) TYPE-2 OU, and (c) no OU for the base case with STEP\_AR1 productivity and CONFINED spatial coherence among stocks in their productivity.



The effect of a safety margin depended to some degree on the type of productivity pattern. Under the STEP\_AR1 pattern, there was little or no decrease in conservation-concern index for TYPE-1 OU as the safety margin was increased from 20% to 40% (Fig. 9a). However, for the CONST and AR1 scenarios, the general patterns for TYPE-1 OU were much more similar to those observed for TYPE-2 OU, including a proportional reduction in conservation-concern index as we stepped up the safety margin from 20% to 40% (Supplemental Table S1).<sup>4</sup>

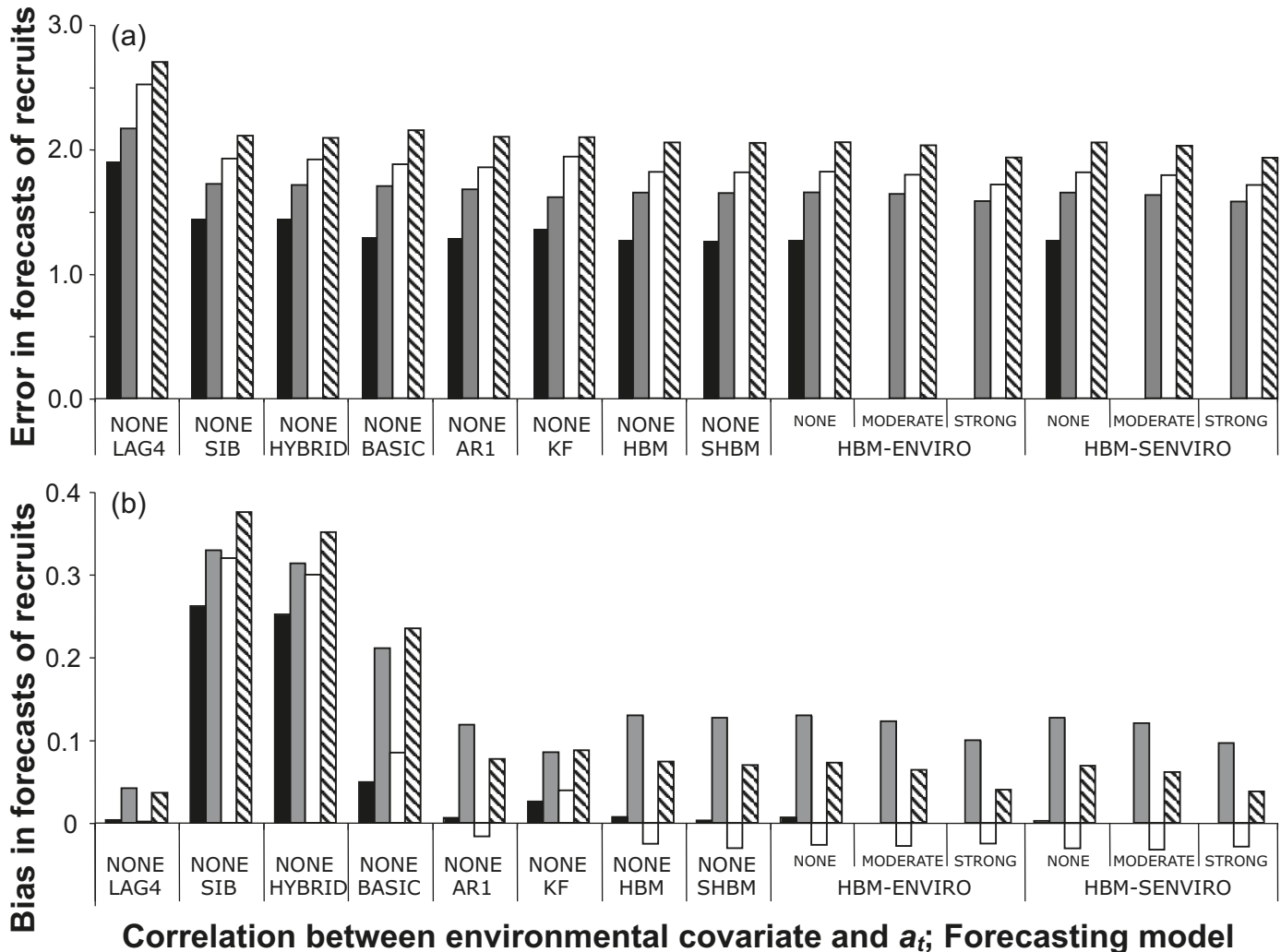
#### Effect of initial conditions

We tested the effect of initial conditions on model performance to ensure that these conditions did not bias our conclusions. In particular, we were concerned that our comparatively high-contrast initial spawner–recruit time series (initial spawner abundances were allowed to vary between 20% and 80% of unfished equilibrium abundance) might have down-played differences in performance among the different stock assessment and forecasting models because these models were all given comparatively high-quality data

to work with. To test this hypothesis, we ran separate simulation trials that used three alternative lower-contrast scenarios (as compared with our base case) in the 20-year initialization period: (1) a constant initial spawner abundance of 50% of the unfished equilibrium and a harvest rate of 0.5, (2) the same as scenario 1 except with a harvest rate of 0.8 to hold the population at low spawner abundance, and (3) the same as scenario 1 except with a harvest rate of zero, which led to a high relatively constant abundance. We found that the effect of conditions in these initial 20 years on performance indicators of a given model were erased within about 30 years of the subsequent 100 simulated years; those 100 years were when we calculated our performance measures. This masking of initial conditions occurred fastest and most thoroughly in scenarios with TYPE-2 OU, but it still occurred even in cases without any OU.

This lack of lasting effect of initial conditions appears to be due to the feedback nature of the system. When initial conditions produce high contrast in spawner abundance, parameter estimates are relatively good compared with low-contrast cases and so the simulated management decisions

**Fig. 10.** Effect of type of climate-induced variability in productivity on (a) error in forecasts of adult recruits (RMSE) and (b) bias (forecast minus actual run size) in those forecasts, given CONFINED spatial coherence, TYPE-2 OU, and a safety margin of zero. The four types of climate-induced variability in productivity were constant productivity (solid bars), step function (shaded bars), autoregressive-lag-1-year (AR(1)) process (open bars), and step function plus AR(1) process (hatched bars). Because it is impossible to create a covariate for a constant time series of  $a_t$ , scenarios with environmental covariates for the hierarchical Bayesian models could not be run with CONST productivity; hence, those cases are left blank.



tend to better achieve escapement goals, thereby leading to lower future contrast in spawner abundance. However, with low-contrast initial situations, parameter estimates are relatively poor and actual escapements tend to vary greatly around the escapement goals, which subsequently leads to high-contrast situations, which in turn gives better parameter estimates, and so on. Thus, the performance measures that we report here for different models largely reflect the dynamics of the internally generated feedbacks in the system, not the initial conditions.

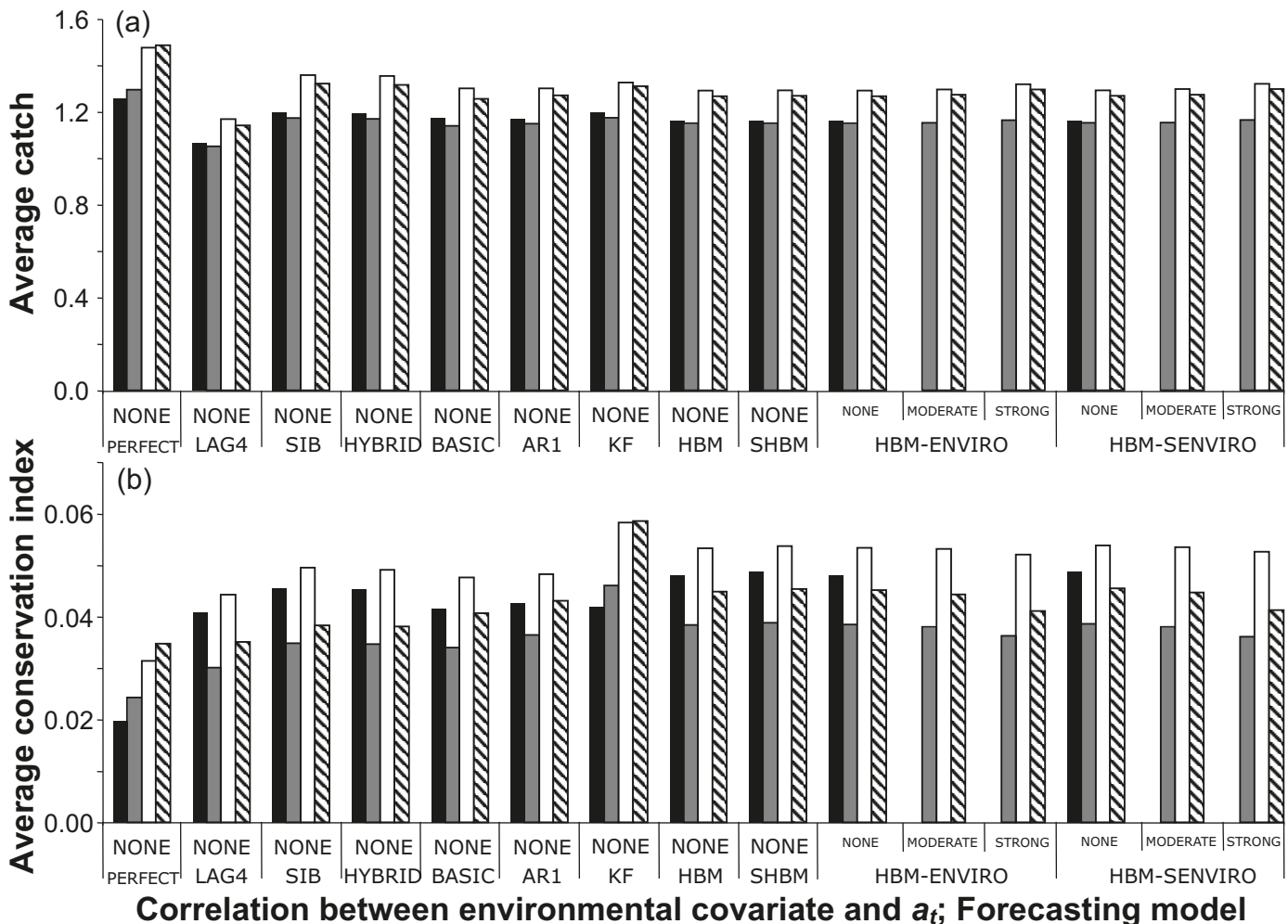
## Discussion

Our results highlight the largely overlooked importance of outcome uncertainty (OU). The greatest overall change in average catch across our simulated scenarios occurred when OU was added, regardless of its type. This effect was larger than the consequences of changes in spatial coherence in

productivities across salmon stocks, safety margin, and average correlation between the environmental variable and productivity parameters. As well, the average index of conservation concern increased most in the presence of TYPE-2 OU, the type in which both actual abundance of recruits and preseason forecasts affect deviations between actual and target spawner abundances. This dominant effect of OU results from its production of a large variance over time in realized escapement, which tends to mask the effects of changes in target escapement produced at other phases in the fishery system, such as the stock assessment phase.

We therefore share the view of other researchers that before proposed new stock assessment and forecasting models are adopted, their statistical performance must be evaluated via simulations of entire fishery systems that include major sources of uncertainty, including OU. Despite the logic of new types of models, their benefits may be unobservable, even in the long run, due to large sources of variation else-

**Fig. 11.** Effect of climate-induced variability in productivity on (a) average catch and (b) average conservation-concern index, given CONFINED spatial coherence, TYPE-2 OU, and a safety margin of zero. The four types of climate-induced variability in productivity were constant productivity (solid bars), step function (shaded bars), autoregressive-lag-1-year (AR(1)) process (open bars), and step function plus AR(1) process (hatched bars).



where in the fishery system. It is thus only feasible to determine such purported improvements when they are placed in the context of how the resulting information is to be used in the fishery system. The closed-loop (or MSE) approach used here is the most appropriate method for doing such evaluations (Walters 1986; Butterworth and Punt 1999; Sainsbury et al. 2000). Several such simulations have documented the substantial effect of OU (e.g., Butterworth and Punt 1999; Peterman et al. 2000; Kell et al. 2005). The June 2008 proposed National Standard 1 Guidelines of the Magnuson–Stevens Fishery Conservation and Management Reauthorization Act (National Marine Fisheries Service 2008) also specify that stock assessment models should explicitly account for OU. Of course, modelers who develop closed-loop simulations have no way of eliminating one key source of uncertainty: how the “true” environment and salmon populations will change under different climatic scenarios. However, if enough different hypothesized “true” scenarios are simulated by the operating model, then it might be possible to draw robust conclusions from results.

We also found that reducing OU does not always produce more desirable outcomes in terms of conservation concerns.

Although scenarios with TYPE-2 OU consistently produced the highest values for the conservation-concern index, that index for TYPE-1 OU was on average only about half of what we observed for runs without OU. In other words, runs without OU produced more cases of conservation concern than runs that included effects of TYPE-1 OU. This counterintuitive result is explained by a combination of two factors. First, the TYPE-1 OU has a slight bias towards producing higher escapements than the given target; it thus acts in a similar way to a safety margin. Second, we found that forecasting and stock assessment models benefited from the large year-to-year variability in spawner abundances created by introducing OU because this variability provided increased contrast (range) in spawner abundances, which led to better parameter estimates (Walters 1986). Scenarios without OU thus usually resulted in poorer fits than their counterparts with OU. Ironically, this result also means that “successful” management that results in close approximation to a fixed escapement and reduces contrast of spawner abundances will degrade parameter estimation in stock assessments and forecasting, a situation noted before by Walters (1986).

Below OU, the next most important influence on average catch and the conservation index was the size of safety margin (0%, 20%, or 40% added to the initially calculated target escapement). A nonzero safety margin evens out differences between outcomes for different types of OU and thus can reduce risk when little is known about the system. However, the cost of invoking a safety margin is lower average catch. Our analyses show a critically important result, however. Reducing OU helps to minimize this impact of safety margin on catch and also improves the effectiveness of a safety margin at reducing the index of conservation concern.

Although more advanced complex models (such as those with environmental covariates and those that took advantage of spatial coherence in productivities across salmon stocks) outperformed simpler models in most respects, we found that model selection had less impact than we expected on average annual catch and average index of conservation concern. In particular, when we simulated either MODERATE or STRONG correlation between the environmental covariate and productivity of salmon stocks (as opposed to NONE), models that included the environmental covariate (HBM-ENVIRO and HBM-SENVIRO) produced 2.4% higher catches and 9% lower indices of conservation concern and 6% smaller errors and 13%–46% lower biases in forecasts of abundance than analogous models without the environmental index (HBM and SHBM). The representation of spatial covariation in the SHBM and HBM-SENVIRO models also helped to improve forecasting performance compared with their counterparts HBM and HBM-ENVIRO, respectively, but the incremental effect was less evident (<1% difference in catch and conservation indices) than when the environmental covariate was added to a model. However, the more complex models produced estimates of Ricker  $a$  and  $b$  parameters that were biased upward by larger amounts than estimates produced by the two basic models (BASIC Ricker and AR1). The only exception was bias in the  $a$  parameter for the AR1 model, which was about the same as that for more complex models. The comparatively higher bias in  $b$  for the HBMs offset reductions in the conservation-concern index resulting from the HBMs' more accurate forecasts. The more complex models' overestimates in  $a$  resulted in overestimates of run size but also overestimated escapement targets, which compensated for overestimates in run size. OU also attenuated the effects of forecasting bias and error. TYPE-1 OU has a built-in safety margin; it tends to produce escapements higher than the target, thereby offsetting forecasting bias of the models. The effect of TYPE-2 OU on escapements is more heavily influenced by the actual abundance of adult recruits than by the forecasting bias and error. The net effect of these factors is that differences among most of the models explored here have relatively little influence on catches and the conservation-concern index.

Thus, we conclude that simpler salmon models used in a management context may be almost as good as more complex models that have many more parameters to estimate and that require considerably more time and expertise to implement. This conclusion is similar to the one made by Hilborn (1979) and Ludwig and Walters (1985). In Hilborn's (1979) case, a simpler model tended to produce relatively precise parameter estimates that happened to be biased such

that they reduced the chance of overfishing. In our case, the effects of OU, in combination with the BASIC Ricker stock assessment model's negative bias in  $b$ , countered the effect of forecasting bias and, as in Hilborn's (1979) case, helped to reduce the chance of overfishing. Adkison (2009) showed more generally that in models of managed natural resources, more complex models may perform more poorly than simpler models. This situation can result from the well-known trade-off between bias and variance in parameter estimates and their interaction with the managers' loss function.

Compared with OU and the size of the safety margin, the effect of climate-induced variability in productivity on average catch and conservation-concern index was relatively small and was about the same magnitude as that of the choice of stock assessment/forecasting model. This relatively small influence of different productivity patterns is good news. When annual updates of parameter estimates are used to set escapement goals, changes in productivity can be tracked sufficiently well to maintain those performance measures within a few percentage points of their initial values. Although there are still huge uncertainties about our ability to forecast effects of future climatically driven environmental conditions, let alone their net effect on salmon populations, the ability to quickly and reliably update estimates of changing productivity of salmon populations is thus an important alternative way of dealing with uncertainty associated with future climatic change.

Although the temporal pattern of climate-induced variability in productivity was not a key influencing factor in our case, different types of productivity patterns nonetheless produced qualitatively different system behaviours. For instance, for the CONST pattern, an increase in safety margin always resulted in a proportional decrease in conservation-concern index, irrespective of OU type. Nevertheless, under the STEP\_AR1 regime, increasing the safety margin from 20% to 40% was only beneficial for TYPE-2 OU but counterproductive for TYPE-1 OU. Thus, if we had only run our simulations with constant productivity, as many stock assessment scientists do, we would have arrived at qualitatively different conclusions about the benefits of safety margins. Therefore, to fully evaluate the potential benefits of introducing new management strategies or methods, it is important that closed-loop simulations explicitly model not just high-frequency interannual variation in survival rates but also low-frequency climate-induced variability (time trends and rapid "regime shifts" to new mean levels). Note that there was considerable short-term, interannual random variability even in our CONST scenario, so the lesson here is not just that variability matters and needs to be simulated but that long-term temporal patterns of variability are also crucial to include.

Together, these results suggest an important conclusion, at least for the types of Pacific salmon systems modeled here. Management objectives might be better met in the future if considerably more effort were put into reducing the deviation between target escapements and actual outcomes (i.e., OU). This could be done, for example, by creating more intensive in-season monitoring, stronger enforcement of regulations, and disincentives for short-term overfishing. Incorporation of empirical models of OU into management targets to counterbalance any expected biases may also help



to change the characteristics of OU and mitigate its undesirable effects (Holt and Peterman 2008).

We have also shown that forecasting error and (or) bias are not necessarily the most useful measures of model performance from a whole-systems perspective because the relationship between forecasts of abundance and outcomes for catch and the conservation-concern index is not straightforward. That is, using a “better” model from the perspective of forecasting RMSE does not necessarily result in a proportionally higher catch and lower conservation-concern index. This is because bias in parameter estimates (especially the Ricker  $b$ ) also plays an important role in determining outcomes, as do safety margins, type of climate pattern, and OU. This situation is analogous to the warning by Walters and Martell (2004, p. 101) that purely statistical criteria (such as the Akaike information criterion) for selecting the “best” model from some group of models are not necessarily appropriate for choosing the “best” model based on management criteria. The reason is that a set of parameters will likely differ in their effects on performance indicators that are important for management decisions. Low bias but high variance in estimates of one parameter may have quite different effects on such performance indicators compared with another parameter that is estimated with larger bias but greater precision.

### Limitations

Even though we evaluated numerous scenarios in the “operating model”, the generality of our conclusions is still limited to the types of conditions that we explored. For instance, higher average Ricker  $a$  parameter values than we used here will produce higher contrast in spawners (and vice versa with lower  $a$  values), so results for other species or regions with different ranges of  $a$  parameters may have somewhat different results than we have shown here. The dominant role of OU shown here will likely be even larger for pink and chum salmon, which tend to have higher forecasting uncertainty than sockeye salmon on which we based our model (Haeseker et al. 2005, 2008). Primarily due to a lack of suitable data on OU, we did not explore mixed-stock fisheries, which arise where several salmon stocks are harvested simultaneously in the same fishery. Typically, those stocks differ not only in their productivity parameters (as we assumed here) but also in equilibrium abundance (which we did not assume). In such mixed-stock fisheries, managers commonly focus attention on the most abundant stocks and tend to overharvest others. We suspect that if we had simulated mixed-stock fisheries, the index of conservation concern would have been much higher under most scenarios, as it would also have been if we had assumed compensatory mortality in the population model (Liermann and Hilborn 1997). In any case, the primary lesson from our simulations would likely remain the same, that is, we need to characterize OU, explicitly incorporate its effects into stock assessment advice to managers, and reduce it as far as possible to increase (or maintain) catch and reduce conservation concerns.

### Acknowledgments

We thank scientists who participated in a workshop at the Pacific Biological Station in Nanaimo, B.C., for helpful

comments on our initial results. We also thank Milo Adkison and Kendra Holt for their thoughtful and thorough reviews of an earlier draft of this paper. Two anonymous reviewers also made useful suggestions. Funding for this research project was provided by the Natural Sciences and Engineering Research Council of Canada, the Canada Research Chairs Program, and the Gordon and Betty Moore Foundation. We are also grateful to Simon Fraser University and WestGrid for providing the high-performance computing resources required to conduct this research.

### References

- Adkison, M.D. 2009. Drawbacks of complex models in frequentist and Bayesian approaches to natural-resource management. *Ecol. Appl.* **19**(1): 198–205. doi:10.1890/07-1641.1. PMID:19323183.
- Banerjee, S., Carlin, B.P., and Gelfand, A.E. 2004. Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC, Boca Raton, Fla.
- Beamish, R.J. (Editor). 1995. Climate change and northern fish populations. *Can. Spec. Publ. Fish. Aquat. Sci.* No. 121.
- Besag, J., York, J., and Mollié, A. 1991. Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Stat. Math.* **43**: 1–59. doi:10.1007/BF00116466.
- Butterworth, D.S. 2007. Why a management procedure approach? Some positives and negatives. *ICES J. Mar. Sci.* **64**(4): 613–617. doi:10.1093/icesjms/fsm003.
- Butterworth, D.S., and Punt, A.E. 1999. Experiences in the evaluation and implementation of management procedures. *ICES J. Mar. Sci.* **56**(6): 985–998. doi:10.1006/jmsc.1999.0532.
- Cass, A., Folkes, M., and Pestal, G. 2003. Methods for assessing harvest rules for Fraser River sockeye salmon. Pacific Scientific Advice Review Committee Working Paper No. S2003-14. Fisheries and Oceans Canada, Nanaimo, B.C.
- de la Mare, W.K. 1998. Tidier fisheries management requires a new MOP (management-oriented paradigm). *Rev. Fish Biol. Fish.* **8**(3): 349–356. doi:10.1023/A:1008819416162.
- Dorner, B., Peterman, R.M., and Haeseker, S. 2008. Historical trends in productivity of 120 Pacific pink, chum, and sockeye salmon stocks reconstructed by using a Kalman filter. *Can. J. Fish. Aquat. Sci.* **65**(9): 1842–1866. doi:10.1139/F08-094.
- Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 2004. Bayesian data analysis. 2nd ed. Chapman and Hall/CRC, London, U.K.
- Haeseker, S.L., Peterman, R.M., Su, Z., and Wood, C.C. 2005. Retrospective evaluation of preseason forecasting models for pink salmon. *N. Am. J. Fish. Manag.* **25**(3): 897–918. doi:10.1577/M04-085.1.
- Haeseker, S.L., Dorner, B., Peterman, R.M., and Su, Z. 2007. An improved sibling model for forecasting chum salmon and sockeye salmon abundance. *N. Am. J. Fish. Manag.* **27**(2): 634–642. doi:10.1577/M06-094.1.
- Haeseker, S.L., Peterman, R.M., Su, Z., and Wood, C.C. 2008. Retrospective evaluation of pre-season forecasting models for sockeye and chum salmon. *N. Am. J. Fish. Manag.* **28**(1): 12–29. [Plus Appendixes 1 and 2 available online (A1-1-A2-7).] doi:10.1577/M06-287.1.
- Hare, S.R., and Francis, R.C. 1995. Climate change and salmon production in the Northeast Pacific Ocean. In *Climate change and northern fish populations*. Edited by R.J. Beamish. *Can. Spec. Publ. Fish. Aquat. Sci.* No. 121. pp. 357–372.
- Hayes, D.B., Brodziak, J.K.T., and O’Gorman, J.B. 1995. Efficiency and bias of estimators and sampling designs for deter-

- mining length–weight relationships of fish. *Can. J. Fish. Aquat. Sci.* **52**(1): 84–92. doi:10.1139/f95-008.
- Hilborn, R. 1979. Comparison of fisheries control systems that utilize catch and effort data. *J. Fish. Res. Board Can.* **36**: 1477–1489.
- Hilborn, R. 1985. Simplified calculation of optimum spawning stock size from Ricker's stock recruitment curve. *Can. J. Fish. Aquat. Sci.* **42**: 1833–1834. doi:10.1139/f85-230.
- Holt, C.A., and Peterman, R.M. 2006. Missing the target: uncertainties in achieving management goals in fisheries on Fraser River, British Columbia, sockeye salmon (*Oncorhynchus nerka*). *Can. J. Fish. Aquat. Sci.* **63**(12): 2722–2733. doi:10.1139/F06-155.
- Holt, C.A., and 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**(7): 1459–1474. doi:10.1139/F08-053.
- Intergovernmental Panel on Climate Change. 2007. Climate change 2007: synthesis report. A report of the Intergovernmental Panel on Climate Change ([www.ipcc.ch/publications\\_and\\_data/publications\\_ipcc\\_fourth\\_assessment\\_report\\_synthesis\\_report.htm](http://www.ipcc.ch/publications_and_data/publications_ipcc_fourth_assessment_report_synthesis_report.htm)).
- Kell, L.T., Pilling, G.M., Kirkwood, G.P., Pastoors, M., Mesnil, B., Korsbrekke, K., Abaunza, P., Aps, R., Biseau, A., Kunzlik, P., Needle, C., Roel, B.A., and Ulrich-Rescan, C. 2005. An evaluation of the implicit management procedure used for some ICES roundfish stocks. *ICES J. Mar. Sci.* **62**(4): 750–759. doi:10.1016/j.icesjms.2005.01.001.
- Liermann, M., and Hilborn, R. 1997. Depensation in fish stocks: a hierarchic Bayesian meta-analysis. *Can. J. Fish. Aquat. Sci.* **54**(9): 1976–1984. doi:10.1139/cjfas-54-9-1976.
- Linhardt, H., and Zucchini, W. 1986. Model selection. John Wiley, New York.
- Ludwig, D., and Walters, C.J. 1985. Are age-structured models appropriate for catch-effort data? *Can. J. Fish. Aquat. Sci.* **42**(6): 1066–1072. doi:10.1139/f85-132.
- Mueter, F.J., Peterman, R.M., and Pyper, B.J. 2002a. Opposite effects of ocean temperature on survival rates of 120 stocks of Pacific salmon (*Oncorhynchus* spp.) in northern and southern areas. *Can. J. Fish. Aquat. Sci.* **59**(3): 456–463. doi:10.1139/f02-020.
- Mueter, F.J., Ware, D.M., and Peterman, R.M. 2002b. Spatial correlation patterns in coastal environmental variables and survival rates of Pacific Salmon in the Northeast Pacific Ocean. *Fish. Oceanogr.* **11**(4): 205–218. doi:10.1046/j.1365-2419.2002.00192.x.
- Mueter, F.J., Pyper, B.J., and Peterman, R.M. 2005. Relationships between coastal ocean conditions and survival rates of salmon in the Northeast Pacific Ocean at multiple lags. *Trans. Am. Fish. Soc.* **134**(1): 105–119. doi:10.1577/T-04-033.1.
- Myers, R.A. 2001. Stock and recruitment: generalizations about maximum reproductive rate, density dependence, and variability using meta-analytic approaches. *ICES J. Mar. Sci.* **58**(5): 937–951. doi:10.1006/jmsc.2001.1109.
- National Marine Fisheries Service. 2008. Magnuson–Stevens Act Provisions; Annual Catch Limits; National Standard Guidelines. Proposed Rule; Request for Comments. *U.S. Fed. Regist.* **73**(111): 32526–32547.
- Peterman, R.M. 1982. Model of salmon age structure and its use in preseason forecasting and studies of marine survival. *Can. J. Fish. Aquat. Sci.* **39**(11): 1444–1452. doi:10.1139/f82-195.
- Peterman, R.M., Pyper, B.J., Lapointe, M.F., Adkison, M.D., and Walters, C.J. 1998. Patterns of covariation in survival rates of British Columbian and Alaskan sockeye salmon (*Oncorhynchus nerka*) stocks. *Can. J. Fish. Aquat. Sci.* **55**(11): 2503–2517. doi:10.1139/cjfas-55-11-2503.
- Peterman, R.M., Pyper, B.J., and Grout, J.A. 2000. Comparison of parameter estimation methods for detecting climate-induced changes in productivity of Pacific salmon (*Oncorhynchus* spp.). *Can. J. Fish. Aquat. Sci.* **57**(1): 181–191. doi:10.1139/cjfas-57-1-181.
- Peterman, R.M., Pyper, B.J., Mueter, F.J., Haeseker, S.L., Su, Z., and Dorner, B. 2009. Statistical models of Pacific salmon that include environmental variables. In *Pacific salmon environment and life history models: advancing science for sustainable salmon in the future*. Edited by E.E. Knudsen and H.J. Michael, Jr. American Fisheries Society Symposium 71, American Fisheries Society, Bethesda, Md.
- Pyper, B.J., Mueter, F.J., Peterman, R.M., Blackburn, D.J., and Wood, C.C. 2001. Spatial covariation in survival rates of Northeast Pacific pink salmon (*Oncorhynchus gorbuscha*). *Can. J. Fish. Aquat. Sci.* **58**(8): 1501–1515. doi:10.1139/cjfas-58-8-1501.
- Pyper, B.J., Mueter, F.J., Peterman, R.M., Blackburn, D.J., and Wood, C.C. 2002. Spatial covariation in survival rates of Northeast Pacific chum salmon. *Trans. Am. Fish. Soc.* **131**(3): 343–363. doi:10.1577/1548-8659(2002)131<0343:SCISRO>2.0.CO;2.
- Pyper, B.J., Mueter, F.J., and Peterman, R.M. 2005. Across-species comparisons of spatial scales of environmental effects on survival rates of Northeast Pacific salmon. *Trans. Am. Fish. Soc.* **134**(1): 86–104. doi:10.1577/T04-034.1.
- Ricker, W.E. 1975. Computation and interpretation of biological statistics of fish populations. *Fish. Res. Board. Can. Bull.* No. 191.
- Rosenberg, A.A., and Brault, S. 1993. Choosing a management strategy for stock rebuilding when control is uncertain. *Can. Spec. Publ. Fish. Aquat. Sci.* No. 120. pp. 243–249.
- Sainsbury, K. 1998. Living marine resources assessment for the 21st century: what will be needed and how will it be provided? In *Fishery stock assessment models*. Edited by F. Funk, T.J. Quinn, J.N. Heifetz, J.E. Ianelli, J.F. Powers, J.F. Schweigert, P.J. Sullivan, and C.-I. Zhang. Alaska Sea Grant College Program, AK-SG-98-01, University of Alaska, Fairbanks, Alaska. pp. 1–40.
- Sainsbury, K.J., Punt, A.E., and Smith, A.D.M. 2000. Design of operational management strategies for achieving fishery ecosystem objectives. *ICES J. Mar. Sci.* **57**(3): 731–741. doi:10.1006/jmsc.2000.0737.
- Schnute, J.T., and Richards, L.J. 1995. The influence of error on population estimates from catch–age models. *Can. J. Fish. Aquat. Sci.* **52**(10): 2063–2077. doi:10.1139/f95-800.
- Smith, A.D.M., Sainsbury, K.J., and Stevens, R.A. 1999. Implementing effective fisheries management systems — management strategy evaluation and the Australian partnership approach. *ICES J. Mar. Sci.* **56**(6): 967–979. doi:10.1006/jmsc.1999.0540.
- Su, Z., Peterman, R.M., and Haeseker, S.L. 2004. Spatial hierarchical Bayesian models for stock–recruitment analysis of pink salmon (*Oncorhynchus gorbuscha*). *Can. J. Fish. Aquat. Sci.* **61**(12): 2471–2486. doi:10.1139/f04-168.
- Thompson, K.R., and Page, F.H. 1989. Detecting synchrony of recruitment using short, autocorrelated time series. *Can. J. Fish. Aquat. Sci.* **46**(10): 1831–1838. doi:10.1139/f89-230.
- Walters, C.J. 1986. Adaptive management of renewable resources. MacMillan, New York.
- Walters, C.J., and Martell, S.J.D. 2004. Fisheries ecology and management. Princeton University Press, Princeton, N.J.
- Walters, C.J., and Parma, A.A. 1996. Fixed exploitation rate strate-

- gies for coping with effects of climate change. *Can. J. Fish. Aquat. Sci.* **53**(1): 148–158. doi:10.1139/cjfas-53-1-148.
- Wood, C.C., Rutherford, D.T., Peacock, D., Cox-Rogers, S., and Jantz, L. 1997. Assessment of recruitment forecasting methods for major sockeye and pink salmon stocks in northern British Columbia. *Can. Tech. Rep. Fish. Aquat. Sci.* No. 2187.

Copyright of Canadian Journal of Fisheries & Aquatic Sciences is the property of NRC Research Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.