North American Journal of Fisheries Management 28:12–29, 2008

DOI: 10.1577/M06-287.1

[Article]

Retrospective Evaluation of Preseason Forecasting Models for Sockeye and Chum Salmon

STEVEN L. HAESEKER,*[1] RANDALL M. PETERMAN, AND ZHENMING SU[2]
School of Resource and Environmental Management, Simon Fraser University,
8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada

CHRIS C. WOOD
Fisheries and Oceans Canada, Pacific Biological Station,
3190 Hammond Bay Road, Nanaimo, British Columbia V9T 6N7, Canada

Abstract.—Using comprehensive data sets for chum salmon *Oncorhynchus keta* (40 stocks) and sockeye salmon *O. nerka* (37 stocks) throughout their North American ranges, we compared the retrospective performance of 11 models in preseason forecasting of adult abundance. Chum and sockeye salmon have more complicated age structures than pink salmon *O. gorbuscha*, which we investigated previously (Haeseker et al. 2005), and this complexity presents new challenges as well as opportunities for forecasting. We extended our previous work to include two new forecasting models that make use of leading indicators: either the survival rate of earlier-maturing pink salmon from the same brood year (Ricker pink salmon index model) or the abundance of earlier-maturing siblings (hybrid sibling model, a new version of the standard sibling model). No single forecasting model was consistently the best for either chum or sockeye salmon, but the hybrid sibling model frequently performed best based on mean absolute error, mean percent error, and root mean square error. As was observed for pink salmon, several naïve models (i.e., simple time series models without explicitly modeled mechanisms) also performed well, as did forecast averaging models composed of two models with the least-correlated forecasting errors. In general, model ranking depended on the particular stock and performance measure used. However, even the top-ranked model for each stock explained on average only 21% of the observed interannual variation in chum salmon recruitment and only 36% of the variation in sockeye salmon recruitment. Although improvements may be possible for some stocks in specific circumstances, a major breakthrough in general forecasting ability seems unlikely given the breadth of stocks and models examined to date. Therefore, better in-season updates and adjustments to fishing regulations and a cautious approach to opening and closing fisheries should remain high priorities.

Prior to the start of each annual fishing season, management agencies on the West Coast of North America forecast the abundances of salmon *Oncorhynchus* spp. by region to provide input to planning by fisheries managers, the fishing industry, and First Nations (e.g., Alaska Department of Fish and Game 2006; Cass et al. 2006; Washington Department of Fish and Wildlife 2006). Generally, these preseason forecasts have wide confidence intervals, and the actual abundances of adult recruits routinely differ from mean or median forecasts—often by large amounts—because of unpredictable variations in factors such as marine survival rates (Fried and Yuen 1987; Cross and Gray 1999; Adkison 2002). Although management agencies subsequently use in-season updates of abundance estimates to adjust management regulations, such differences between actual recruits and preseason forecasts usually result in missed management targets, reduced economic benefits, or increased conservation concerns (Bocking and Peterman 1988; Holt and Peterman 2006).

For many years, scientists have developed and applied a wide variety of forecasting models to improve on this situation. Models vary from simple moving-average time series models (Wood et al. 1997) to complex stock–recruitment models that include environmental variables (Adkison et al. 1996) or even non-mechanistic models based on neural networks (Zhou 2003). These models have had varying, but generally poor success at forecasting past abundances (Fried and Yuen 1987; Adkison and Peterman 2000; Cass et al. 2006).

One problem that may have hindered development of better forecasting models is the lack of extensive testing of each model across a wide range of populations and regions. When performance has been

---

* Corresponding author: steve_haeseker@fws.gov

[1] Present address: U.S. Fish and Wildlife Service, Columbia River Fishery Program Office, 1211 Southeast Cardinal Court, Suite 100, Vancouver, Washington 98683, USA.

[2] Present address: Institute for Fisheries Research, 1109 North University Avenue, 212 Museums Annex Building, Ann Arbor, Michigan 48109-1084, USA.

Received December 15, 2006; accepted March 26, 2007

compared among multiple models, it has usually been within particular regions and species, such as northern or southern British Columbia (BC) sockeye salmon *O. nerka* (Wood et al. 1997; Cass et al. 2006) or southeast Alaska pink salmon *O. gorbuscha* (Adkison 2002). An exception to these regionally based studies is our previous work (Haeseker et al. 2005), which evaluated the forecasting ability of eight preseason forecasting models for 43 pink salmon stocks from Washington State, BC, and Alaska over a total of 783 stock-years. We (Haeseker et al. 2005) evaluated those eight models for each pink salmon stock by means of retrospective analysis, a type of cross validation (Shao 1993) that uses historical data up to a given year to fit various forecasting models, makes forecasts of the next year's recruitment, compares forecasts with the actual observed recruitment, adds that new actual data point into the database, re-fits the models, makes forecasts for the next year, and so on through the entire data set. The performance of each model is then summarized across years.

We found that no single forecasting model was consistently the best performer across the 43 pink salmon stocks (Haeseker et al. 2005). However, we found that two naïve time series models (i.e. those without explicitly modeled mechanisms) performed best more frequently than other models, although there were specific populations for which other, more complex models made better forecasts (Haeseker et al. 2005). For instance, in 21% of the populations, the best forecasts were made by a multistock mixed-effects (ME) stock–recruitment model, which included early summer sea surface temperature (SST) as an independent variable along with spawner abundance (Mueter et al. 2002a).

Here we extend our retrospective evaluation of a variety of forecasting models for pink salmon populations by applying a similar approach to evaluating forecasting models for 40 populations of chum salmon *O. keta* and 37 populations of sockeye salmon. We sought to determine (1) whether particular models might be more widely beneficial across chum and sockeye populations than in our pink salmon analysis, (2) whether overall average forecasting errors would be larger or smaller for chum and sockeye salmon than for pink salmon, and (3) whether the multiple ages at maturity for chum and sockeye salmon help or hinder forecasting performance. Several forecasting models that we investigated for chum and sockeye salmon were the same as or similar to those used for pink salmon (Haeseker et al. 2005), but a few models were new because of differences in age structure between the species. Unlike pink salmon, which all mature at age 2, chum and sockeye salmon typically mature and return to their natal spawning areas as recruits of age 3–6 depending on the region (Groot and Margolis 1991). Spawners of all three species die after spawning. Complexity in age structure presents new challenges and opportunities for forecasting that do not arise in the forecasting of pink salmon populations. Challenges exist owing to interannual variability in age structure, and opportunities arise through using abundances of younger-aged fish as leading indicators of a subsequent year's abundance.

We evaluated two chum and sockeye salmon forecasting models that used such leading indicators. One incorporated indices of early ocean survival based on nearby pink salmon populations that entered the ocean during the same year as chum salmon fry and sockeye salmon smolts (Anderson and Bailey 1974). The other was a variant on what we call a standard sibling model, which describes the relationship between returns of successive age-classes produced in the same brood year (Alaska Department of Fish and Game 1981; Peterman 1982). These and other models are detailed below. Altogether, we compared the efficacy of 11 types of forecasting models for chum and sockeye salmon.

Methods

Data.—We compiled data on the abundance of spawners and recruits for 37 sockeye salmon stocks and 40 chum salmon stocks in the northeastern Pacific from northwestern Alaska to southwestern Washington (USA) and including BC (Tables 1, 2). The stocks in this analysis encompass nearly all of the major sockeye and chum salmon stocks managed in the Northeast Pacific. Spawner abundance (S) included both sexes; recruits (R) included S and total catch of both sexes. Time series of S and R ranged in duration from 22 to 47 years (average = 40 years) for sockeye salmon stocks and from 15 to 35 years (average = 26 years) for chum salmon stocks. Further details on data sources and their compilation can be found in Peterman et al. (1998), Mueter et al. (2002b), and Pyper et al. (2002). The age composition data for adult recruits were sufficient to develop sibling models for all 37 sockeye salmon stocks but only 21 of the 40 chum salmon stocks (Table 2). For each of the remaining 19 chum salmon stocks, the age composition was assumed to be identical over time to the average age composition of the nearest neighboring chum salmon stock.

Two of the models used early summer SST as an explanatory variable because it helps explain some of the variation in survival rates of pink, chum, and sockeye salmon (Mueter et al. 2002a). The SST data were compiled from the Comprehensive Ocean-Atmosphere Data Set (COADS; www.cdc.noaa.gov/coads)

TABLE 1.—Summary of data sets for 37 wild sockeye salmon stocks used in an analysis of adult abundance forecasting models (N = the number of complete brood years available). This information was adapted from Peterman et al. (1998) and Mueter et al. (2002b).

Region	Stock number	Stock or district	Brood years	N	Source[a]
Washington	1	Lake Washington	1967–1993	27	1
Fraser River, British Columbia	2	Adams	1950–1996	47	2
	3	Birkenhead	1950–1996	47	2
	4	Bowron	1950–1996	47	2
	5	Chilko	1950–1996	47	2
	6	Cultus	1950–1996	47	2
	7	Gates	1952–1996	45	2
	8	Horsefly	1950–1996	47	2
	9	Nadina	1950–1996	47	2
	10	Pitt	1950–1996	47	2
	11	Portage	1950–1996	47	2
	12	Raft	1950–1996	47	2
	13	Seymour	1950–1996	47	2
	14	Stellako	1950–1996	47	2
	15	Early Stuart	1950–1996	47	2
	16	Late Stuart	1950–1996	47	2
	17	Weaver	1950–1996	47	2
Central British Columbia	18	Long Lake	1973–1994	22	3
Northern British Columbia	19	Skeena	1959–1994	26	4
	20	Nass	1967–1991	25	4
Central Alaska	21	Copper	1961–1993	33	5
	22	Cook	1968–1992	25	6
Southwestern Alaska	23	Ayakulik	1965–1993	29	7
	24	Frazer	1965–1993	29	7
	25	Early Upper Station	1969–1993	25	7
	26	Late Upper Station	1970–1993	24	7
	27	Black	1950–1993	44	7
	28	Chignik	1950–1993	44	7
Bristol Bay, Alaska	29	Branch	1956–1995	40	8
	30	Egegik	1956–1995	40	8
	31	Igushik	1956–1995	40	8
	32	Kvichak	1956–1995	40	8
	33	Naknek	1956–1995	40	8
	34	Nuyakuk	1956–1995	40	8
	35	Togiak	1956–1995	40	8
	36	Ugashik	1956–1995	40	8
	37	Wood	1956–1995	40	8

[a] Sources are personal communications from (1) Jeff Haymes, Washington Department of Fish and Wildlife, Montesano; (2) Jim Woodey, Pacific Salmon Commission, Vancouver, BC; (3) Chris Wood, Canada Department of Fisheries and Oceans (CDFO), Nanaimo, BC; (4) L. Jantz, CDFO, Prince George, BC; (5). John Wilcock, Alaska Department of Fish and Game (ADFG), Cordova; (6) David Waltemyer, ADFG, Soldotna; (7) Patti Nelson, ADFG, Kodiak; and (8) Bev Cross, ADFG, Anchorage.

and were summarized using the methods described by Mueter et al. (2002b). The SST time series used in our forecasting models represent the SST anomalies (°C) from long-term means at each location corresponding to the coastal areas occupied by each sockeye or chum salmon stock during early ocean residence (April–July for stocks in Washington, BC, and southeastern Alaska; May–August for most Alaska stocks; and June–September for western Alaska stocks; further details are given in Mueter et al. 2002a).

We also examined a model for forecasting sockeye and chum salmon recruitment that relied partly on data for pink salmon from the same region. Previous work showed statistically significant ($P < 0.05$), positive correlations between marine survival rates of pink and chum salmon from the same brood year in southern BC (e.g., Anderson and Bailey 1974; Peterman 1987). In their broader analyses of more than 110 populations from Washington, BC, and Alaska, Pyper et al. (2005) found that region-specific survival rate indices of pink, chum, and sockeye salmon stocks were positively correlated when cohorts of juveniles had the same year of ocean entry; pink and chum salmon survival rates were more strongly correlated than those of pink and sockeye salmon. If shared environmental conditions cause survival rates to be positively correlated between species, then survival rates for earlier-returning age-2 pink salmon from a given brood year could be used to forecast survival rates, and hence recruitment, of later-returning age-3 or older (age-3+) sockeye and chum salmon from the same brood year.

To evaluate the forecasting utility of this positive

TABLE 2.—Summary of data sets for 40 wild chum salmon stocks used in an analysis of adult abundance forecasting models (N = the number of complete brood years available). Asterisks indicate stocks with sufficient age composition data for sibling models. This information was adapted from Pyper et al. (2002).

Region	Stock number	Stock or district	Brood years	N	Source[a]
Washington	1	Willapa Bay*	1968–1993	26	1
	2	Grays Harbor*	1969–1993	25	1
Puget Sound, Washington	3	Skagit River*	1968–1993	26	2
	4	Nooksack–Somish*	1968–1993	26	2
	5	Stillaguamish–Snohomish*	1968–1993	26	2
	6	Hood Canal*	1968–1993	26	2
	7	South Sound (early)*	1968–1993	26	2
	8	South Sound (fall)*	1968–1993	26	2
	9	South Sound (winter)*	1968–1993	26	2
Southern British Columbia	10	Fraser (all)*	1959–1992	34	3
	11	Inner South Coast less Fraser River*	1959–1992	34	3
Central British Columbia	12	Area 10	1960–1994	35	4
	13	Area 9	1960–1994	35	4
	14	Area 8*	1960–1994	35	4
Northern British Columbia	15	Area 6*	1965–1992	28	5
Southeastern Alaska	16	Alsek[b]	1961–1986	26	6
	17	Kadashan	1969–1983	15	7
Central Alaska	18	Prince William Sound*	1966–1993	28	8
	19	Cook Inlet (outer)[c]	1964–1993	30	9
	20	Cook Inlet (south)[d]	1964–1993	30	9
	21	Cook Inlet (upper)	1972–1993	22	10
	22	Kamishak[e]	1968–1993	22	9
Southwestern Alaska	23	Kodiak	1962–1991	20	11
	24	Chignik (central)	1962–1991	20	12, 13
	25	Chignik (eastern)	1962–1991	20	12, 13
	26	Chignik (western)	1962–1991	20	12, 13
	27	Perryville	1962–1991	20	12, 13
Alaska Peninsula	28	Northern	1962–1993	32	12, 14
	29	Izembek–Moffet	1962–1993	32	12, 14
	30	Bechevin–Swanson	1962–1993	32	12, 14
	31	Southeast–South-central	1962–1993	32	12, 15
	32	Southwest–Unimak	1962–1993	32	12, 15
Bristol Bay, Alaska	33	Nushagak*	1974–1993	20	16
	34	Togiak*	1978–1992	15	16
Norton Sound, Alaska	35	Yukon (fall)[f]*	1974–1995	22	17, 18
	36	Anvik*	1972–1993	22	17, 19
	37	Andreafsky*	1972–1995	24	17, 20
	38	Kwiniuk–Tubutulik*	1965–1995	31	17, 21
	39	Norton Sound District 1	1976–1994	19	17, 22
	40	Kotzebue Sound*	1962–1979	18	23

[a] Sources are (personal communication unless otherwise noted) as follows: (1) Rick Brix and John Linth, Washington Department of Fish and Wildlife (WDFW), Montesano; (2) Jeff Haymes, WDFW, Olympia; (3) Ryall et al. (1999); (4) Ron Goruk, Canada Department of Fisheries and Oceans (CDFO), Prince Rupert, BC; (5) Les Jantz, CDFO, Prince Rupert; (6) Tim Zadina, Alaska Department of Fish and Game (ADFG), Ketchikan; (7) Ben Van Alen, ADFG, Douglas; (8) Mark Willette, ADFG, Soldotna; (9) Ted Otis, ADFG, Homer; (10) Ken Tarbox, ADFG, Soldotna; (11) Kevin Brennan, ADFG, Kodiak; (12) Denby Lloyd, ADFG, Kodiak; (13) Owen and Sarafin (1999); (14) Murphy et al. (1999); (15) Campbell et al. (1999); (16) Michael Link, ADFG, Anchorage; (17) Doug Eggers, ADFG, Juneau; (18) Eggers (2001); (19) Clark and Sandone (2001); (20) Clark (2001a); (21) Clark (2001b); (22) Clark (2001c); and (23) Bigler (1985).

[b] Spawner abundances for the East Alsek River were used as an index of escapement; catch data were total harvests in Districts 181–189.

[c] Sum of Port Graham and Dogfish Lagoon data sets.

[d] Sum of Port Dick and Rocky River data sets.

[e] Sum of Bruin, McNeil, Big Kamishak and Little Kamishak rivers, Ursus Cove and Lagoon, and Cottonwood Creek–Iniskin River data sets.

[f] Sum of fall run chum for the Tanana River, Yukon River main stem, and Yukon River tributary data sets.

association between pink salmon and either sockeye or chum salmon, we added an index of pink salmon survival rate as another independent variable (besides spawners) to a Ricker stock–recruit model for forecasting chum and sockeye salmon recruitment. This annual index was derived as follows. For each of the 43 pink salmon populations from Washington to Alaska, we fit a standard Ricker model to the pink salmon spawner–recruit data and calculated time series of annual $\log_e(R/S)$ residuals from that model, producing an index of brood year survival that removed within-stock, density-dependent effects. Pyper et al. (2001:

their Figure 5) identified 14 regions in which these spawner-to-recruit residuals, or survival rate indices, were significantly and positively correlated among stocks. The time series of annual stock-specific residuals were then standardized to have a mean of 0 and a standard deviation of 1. We then averaged these standardized values across stocks within each of the 14 regions to provide an annual index of pink salmon survival for that region. Finally, we assigned each chum and sockeye salmon stock to 1 of the 14 regions based on the location of ocean entry. However, the Willapa Bay and Grays Harbor (Washington) chum salmon stocks (Table 2) did not have pink salmon stocks nearby, so we were unable to develop pink salmon survival indices for these two stocks. To standardize comparisons across species with different freshwater life histories, we designated survival indices by year of ocean entry. Pink and chum salmon both enter the ocean as age-0 juveniles (fry in brood year $t+1$), whereas sockeye salmon typically enter the ocean as age-1 or age-2 juveniles (smolts in $t+2$ or $t+3$). For sockeye salmon stocks with mixed juvenile ages, we weighted the pink salmon survival indices by the stock-specific average proportions of juvenile sockeye salmon entering the ocean at ages 1 and 2.

Models.—We examined a minimum set of models representing the major types in common use, and we added a few more that seemed promising. First, we used three naïve time series models to forecast recruits. The term “naïve” indicates that the models do not explicitly include spawner abundance or environmental variables as independent variables and do not require statistical parameter estimation. They merely summarize recent information on adult recruits (returns), which mature and return to freshwater 3–5 years after being spawned. The first naïve model, $R(\mathrm{yr}-1)$, is

$$R_{\mathrm{yr}} = R_{\mathrm{yr}-1} + \varepsilon_{\mathrm{yr}}, \tag{1}$$

where R_{yr} is the forecasted return for year yr, $R_{\mathrm{yr}-1}$ is the observed return during the previous year (yr − 1), and $\varepsilon_{\mathrm{yr}}$ is the residual error, where $\varepsilon_{\mathrm{yr}} \sim N(0, \sigma^2)$. The second naïve model, $R(\mathrm{yr}-4)$, is

$$R_{\mathrm{yr}} = R_{\mathrm{yr}-4} + \varepsilon_{\mathrm{yr}}, \tag{2}$$

where $R_{\mathrm{yr}-4}$ is the return 4 years prior to the forecasted return. This model uses the observed recruitment 4 years prior as the forecast for the current year, as might seem appropriate in stocks that exhibit 4-year cycles in abundance. Compared with the other candidate naïve models, such as an $R(\mathrm{yr}-3)$ or $R(\mathrm{yr}-5)$ model, the $R(\mathrm{yr}-4)$ naïve model has demonstrated lower root mean square error (RMSE) of the forecasts for sockeye and chum salmon stocks, in which fish most commonly mature at age 4 (Haeseker et al. 2007). The third naïve model (4-year average) is

$$R_{\mathrm{yr}} = \frac{\sum_{t=1}^{4} R_{\mathrm{yr}-t}}{4} + \varepsilon_{\mathrm{yr}}, \tag{3}$$

where $R_{\mathrm{yr}-t}$ is the observed return during year yr − t ($t = 1, \ldots, 4$). This model uses a 4-year moving average as the forecast for the current year. While recruitment errors in stock–recruitment models are generally believed to be multiplicatively lognormal (Peterman 1981), these naïve models may have a different error structure because they utilize simple summaries of previous recruitment to forecast future recruitment. In support of our assumption that recruitment errors for the naïve models were additive and normal, we found that the normality assumption for the errors held in the majority (71%) of the stocks based on Kolmogorov–Smirnov tests for normality.

The equations and terms pertaining to 5 of the 11 models that we evaluated (Table 3) have been described previously (Peterman et al. 2000; Mueter et al. 2002a; Haeseker et al. 2005). Three of these were linearized versions of the single-stock Ricker stock–recruitment function to forecast recruits (Ricker; Ricker autoregressive, AR[1]; and Ricker SST). We also evaluated a Kalman filter (KF) estimation method of the Ricker stock–recruitment model (Peterman et al. 2000; Haeseker et al. 2005), which allowed for temporal changes in productivity (i.e., the Ricker a parameter). Capitalizing on the recent work of Mueter et al. (2002a), we evaluated the forecasting performance of a hierarchical, multistock, ME Ricker stock–recruitment model that incorporated SST. This hierarchical model takes advantage of the observed similarity among nearby stocks' interannual variation in survival rate, which arises from sharing similar environments (Peterman et al. 1998; Pyper et al. 2001, 2002; Su et al. 2004). We considered this type of multistock model because it results in smaller bias and greater precision in parameter estimates than a model that is fit to data for a single stock (Su et al. 2004). The ME model was fit to chum and sockeye salmon stocks separately.

Of the new models we evaluated that are not listed in Table 3, one was a Ricker model that incorporated an index of pink salmon survival, namely,

$$\log_e(R_t/S_t) = a - bS_t + \delta \mathrm{PI}_{t+k} + \varepsilon_t, \tag{4}$$

where δ is the estimated effect of the pink salmon survival index on $\log_e(R_t/S_t)$; PI_{t+k} is the pink salmon survival index for the region of covariation encompassing the ocean entry point of the stock in year $t+k$, when juvenile chum or sockeye salmon from brood

TABLE 3.—Equations for 5 of the 11 forecasting models of adult chum and sockeye salmon abundance evaluated in this study. Symbols are as follows: S_t is the abundance of spawners in brood year t; R_t is the total abundance of adult recruits produced by the S_t spawners; a, a_t, a_i, b, and b_i are parameters of the basic Ricker model components; SST is early-summer sea surface temperature; γ is the fixed effect and g_i the random effect of SST on $\log_e(R/S)$; α is a fixed intercept describing productivity common to all salmon stocks of a given species and area (northern or southern stocks, as described by Mueter et al. 2002a); ϕ is the first-order autocorrelation coefficient; $\varepsilon_t \sim N(0, \sigma^2)$; and $\upsilon_t \sim N(0, \sigma_t{}^2)$.

Model	Equation
Ricker[a]	$\log_e(R_t/S_t) = a - bS + \varepsilon_t$
Ricker AR(1)[a]	$\log_e(R_t/S_t) = a - bS_t + \varepsilon_t$, where $\varepsilon_t = \phi\varepsilon_{t-1} + \upsilon_t$
Kalman filter[a,b]	$\log_e(R_t/S_t) = a_t - bS_t + \varepsilon_t$, where $a_t = a_{t-1} + \upsilon_t$
Ricker SST[a]	$\log_e(R_t/S_t) = a - bS_t + \gamma \mathrm{SST}_{t+k} + \varepsilon_t$
Mixed effects[a,c]	$\log_e(R_{it}/S_{it}) = \alpha + a_i - b_iS_{it} + \gamma \mathrm{SST}_{i,t+k} + g_i\mathrm{SST}_{i,t+k} + \varepsilon_{it}$, where $\varepsilon_{it} = \phi\varepsilon_{it-1} + \upsilon_t$

[a] Haeseker et al. (2005).

[b] Peterman et al. (2000).

[c] Mueter et al. (2002a).

year t migrated to the ocean; a and b are estimated parameters; and $\varepsilon_t \sim N(0, \sigma^2_\varepsilon)$. For chum salmon, k equals 1. In the case of sockeye salmon, k is 2 or 3; the pink salmon survival index values were weighted according to the proportion of sockeye salmon smolts from brood year t that migrated to the ocean in year $t + 2$ as opposed to $t + 3$. This model is referred to as the Ricker pink salmon index model. Later, we use the term "stock–recruitment-type models" to refer to the group that includes the Ricker, Ricker SST, Ricker AR(1), KF, ME, and Ricker pink salmon index models.

To generate forecasts of adult recruits produced by each brood year, $\hat{R}_t$, each model except the naïve models required back-transformation to estimate the forecasted mean number of brood year recruits on the arithmetic scale. Accounting for the well-known bias associated with back-transforming lognormally distributed variables (Beauchamp and Olson 1973), forecasts were generated using the equation

$$\hat{R}_t = \exp\left(\hat{y}_t + \frac{\hat{\sigma}^2}{2}\right) \cdot S_t, \qquad (5)$$

where $\hat{R}_t$ is the forecast of the mean number of recruits resulting from brood year t, $\hat{y}_t$ is the estimate of $\log_e(R_t/S_t)$, and $\hat{\sigma}^2$ is the appropriate variance of the residuals; for the Ricker, Ricker SST, and KF models, $\hat{\sigma}^2_\varepsilon$ was used in equation (5). For the Ricker AR(1) model, the residual variance after accounting for autocorrelation ($\hat{\sigma}^2_\upsilon$) was used in equation (5). For the ME model, we applied equation (5) by estimating the within-stock residual variance (i.e., $\hat{\sigma}^2_t$) rather than the overall $\hat{\sigma}^2_\upsilon$, which was estimated based on all stocks combined. Estimates of brood year recruitment ($\hat{R}_t$) were allocated to return years ($\hat{R}_{\mathrm{yr}}$) according to average age composition in each stock. Exploratory analyses revealed that age composition averaged over all brood years provided more accurate forecasts than did shorter moving averages.

We also evaluated the forecasting performance of a new type of sibling model, the hybrid sibling model described by Haeseker et al. (2007). The standard sibling model (Peterman 1982) assumes a linear relation between raw abundances of two sibling groups in log space, that is,

$$\log_e(R_{d,t}) = a + b\log_e(R_{d-1,t-1}) + \varepsilon_t, \qquad (6)$$

where $R_{d,t}$ is the abundance of the later-returning siblings of age d in year t; $R_{d-1,t-1}$ is the abundance of the earlier-returning siblings of age $d - 1$ in year $t - 1$; a and b are estimated parameters; and $\varepsilon_t \sim N(0, \sigma^2_\varepsilon)$. This form of error term is based on the frequently observed multiplicative lognormal variation of salmon marine survival rates (Peterman 1981). The hybrid sibling model differs from the standard sibling model in that it uses the above sibling model equation to generate a forecast of $R_{d,t}$ when the sibling relationship is "strong" (i.e., characterized by a low $\hat{\sigma}^2_\varepsilon$) but uses a naïve $R(t - 4)$ model when the relationship is "weak" (i.e., characterized by a high $\hat{\sigma}^2_\varepsilon$). Thus, the hybrid sibling model switches between the standard sibling model and a naïve $R(t - 4)$ model depending on the residual variance estimate ($\hat{\sigma}^2_\varepsilon$) of the sibling relationships. The threshold $\hat{\sigma}^2_\varepsilon$ values used to determine strong or weak (1.09 for chum salmon, 2.53 for sockeye salmon) and thus which model to use were based on optimization analyses (Haeseker et al. 2007). Haeseker et al. (2007) found that the hybrid sibling model generally outperformed the standard sibling model. Because of those results and in an effort to minimize the number of models evaluated, we only considered the hybrid sibling model. As with the Ricker models, generating forecasts of adult recruits ($\hat{R}^*_{d,t}$) based on equation (6) required back-transformation to estimate the forecasted mean number of recruits on the arithmetic scale using methods analogous to those described for equation (5). When a forecast of

$R_{d,t}$ was generated using the $R(t - 4)$ model, no back-transformation was necessary because the forecast was already on the arithmetic scale.

Preseason salmon forecasts have also been generated by averaging forecasts of several individual models (Fried and Yuen 1987). Forecast averaging has been shown to increase the precision of forecasts through cancellation of random errors (Clemen 1989), but there are no accepted rules for selecting the appropriate sets of models to include in the forecast averaging procedure. Bates and Granger (1969) found that averaging was most effective when errors in the component models were uncorrelated. Little cancellation of random observation errors will occur if the component data series are highly positively correlated because errors are similar in direction and magnitude. Using the same logic, Ridley (1999) suggested that negatively correlated time series would work best for forecast averaging to increase precision of forecasts. Based on these observations, we generated a time series of forecasting errors for each of the 10 individual forecasting models above, estimated the pairwise correlations among these series, and then selected the two models with the lowest correlation for use in a forecast averaging model.

Retrospective analysis.—We utilized retrospective analysis to evaluate the performance of the 11 forecasting models for sockeye and chum salmon. Only data that would have been available to make a forecast for some past year were used for estimating model parameters and generating the forecast. Our retrospective analysis produced a time series of forecasting errors for each model by (1) iteratively stepping forward through time as each step added a new year to the estimation data set, (2) generating a forecast, and (3) comparing the forecast with the observed value. This method produces “out-of-sample” forecasts that provide a rigorous assessment of each forecasting model’s performance, just as if it had been used historically (Shao 1993, Haeseker et al. 2005).

We initialized each of the 11 forecasting models with data from the first 10 brood years for each stock. After accounting for the data needed for model initialization, the period available for forecasting individual chum salmon stocks ranged from 5 to 25 years (average = 17 years); the period available for forecasting individual sockeye salmon stocks ranged from 8 to 37 years (average = 29 years). Altogether, each model was evaluated across 665 stock-years for chum salmon and 1,081 stock-years for sockeye salmon.

Performance measures.—We used four performance measures to characterize the central tendency and variability in the distribution of annual forecasting errors: mean raw error (MRE), mean absolute error (MAE), mean percent error (MPE), and RMSE. Each of these performance measures has been used in the past to evaluate forecasting models for Pacific salmon by characterizing the differences between the forecasted ($\hat{R}_{\text{yr}}$) and observed (R_{yr}) returns in year yr. The raw error (e_i) was calculated as

$$e_i = \hat{R}_{\text{yr}} - R_{\text{yr}}. \tag{7}$$

Positive values for the raw errors represent forecasts that are too high, whereas negative values represent forecasts that are too low. To obtain the MRE, the raw errors were averaged over the number of years (n) that were forecasted within each stock. This MRE thus reflects the overall bias of forecasts. However, large overestimates in some years can be offset to some extent by large underestimates in others, which would tend to produce an MRE close to 0. Thus, to reflect the magnitude of forecasting errors encountered each year regardless of sign, we calculated the MAE using the mean absolute value of raw errors (i.e., $\Sigma\ |e_i|/n$). We also calculated the traditional RMSE, which provides a measure of forecast error variance and can be used to construct confidence intervals for the forecasts, as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(e_i)^2}{n}}. \tag{8}$$

The model that produces the lowest RMSE would also produce the narrowest confidence intervals. We did not calculate MRE, MAE, and RMSE over all stocks and years because results would have been dominated by stocks with the highest abundance. Therefore, we first calculated the percent error for each annual forecast of returns (1 stock-year), that is,

$$\text{Percent error} = \left(\frac{\hat{R}_{\text{yr}} - R_{\text{yr}}}{R_{\text{yr}}}\right) \cdot 100. \tag{9}$$

The MPE was then calculated by averaging the percent errors from equation (9) over the number of years that were forecast within each stock. To evaluate model performance across all stocks and years, we also calculated the overall average MPE for each model and species. However, it should be noted that for the overall MPE, more weight is given to stocks that have longer time series by putting equal weight on each stock-year.

For each stock and performance measure, we ranked the forecasting models from 1 (best) to 11 (worst) based on the absolute value of the results for MRE, MAE, MPE, and RMSE. We also calculated several

TABLE 4.—Correlations among the annual raw errors for 11 adult abundance forecasting models (described in text) for 37 sockeye salmon stocks (R = forecast based on the return 1 or 4 years earlier; SST = sea surface temperature; AR[1] = first-order autoregressive; KF = Kalman filter; and ME = mixed effect).

Model	R(yr − 1)	R(yr − 4)	4-year average	Ricker	Ricker SST	Ricker AR(1)	KF	ME	Hybrid sibling	Pink salmon index
R(yr − 4)	0.20									
4-year average	0.68	0.68								
Ricker	0.10	0.27	0.19							
Ricker SST	0.24	0.47	0.37	0.67						
Ricker AR(1)	0.14	0.33	0.22	0.95	0.79					
KF	0.08	−0.01	−0.01	0.23	0.42	0.36				
ME	0.17	0.44	0.26	0.50	0.88	0.69	0.45			
Hybrid sibling	0.30	0.66	0.56	0.22	0.47	0.30	0.18	0.48		
Pink salmon index	0.07	0.25	0.15	0.99	0.67	0.96	0.25	0.53	0.21	
4-year average + KF	0.42	0.35	0.51	0.29	0.55	0.43	0.85	0.52	0.44	0.29

other similar performance measures, including mean absolute percent error and median versions of the above measures to reflect central tendency, but the rankings for each of these alternative performance measures were well correlated ($r > 0.60$) with at least one of our four main measures. We therefore report results using only the above four indicators.

We purposely avoided use of only a single performance measure for ranking forecasting models because opinions vary among salmon forecasters, managers, and users about which measure is most useful and informative. For instance, MRE and MPE reflect long-term bias in forecasts (i.e., how much a model tends to over- or underestimate recruitment), whereas MAE measures the average magnitude of annual forecasting errors regardless of sign (Chatfield 2000). Forecasters and managers might be more interested in minimizing MRE and MPE, whereas someone in the commercial fishing industry might put more weight on minimizing MAE. The RMSE reflects both bias and precision (Rice 1995), yet it is not universally used among forecasters, let alone other groups. In appendix Tables A.1.1 and A.2.1 (available in the online version of this article at afs.allenpress.com), we report the rank of each forecasting model for each performance measure and population. Readers can then interpret these results based on their preferred performance measures.

For each stock, we also calculated the coefficient of determination (r^2) of the forecasted versus observed returns for the highest-ranking model, where rank was based on the RMSE. The r^2 represents the proportion of variability in recruitment explained by the best forecasting model. By reporting these values for the highest-ranked models, we provide an upper bound for the proportion of variance that can be explained in retrospective analyses among our suite of forecasting models.

Results

For sockeye salmon, we found mainly positive correlations in raw errors among the 10 individual models across stocks and years (Table 4). Correlations were higher among stock–recruitment-type models (average $r = 0.62$) and among naïve-type models (average $r = 0.52$) but lower between different types of models (average $r = 0.21$). The most negative correlation was between the 4-year average model and the KF model ($r = -0.01$), so these models were selected for the composite forecast averaging model for sockeye salmon.

For chum salmon, we also found positive correlations in raw errors among the 10 individual models across all stocks and years (Table 5), but the correlations were generally higher than those for sockeye salmon. Correlations were especially high among stock–recruitment models (all $r \geq 0.71$; average $r = 0.85$) but were also high among the naïve models (average $r = 0.62$) and between the naïve and stock–recruitment models (average $r = 0.59$). All correlations were positive, but the smallest positive correlation was between the R(yr − 1) and Ricker pink salmon index models ($r = 0.29$), so these were selected for the forecast averaging model for chum salmon. For the Willapa Bay and Grays Harbor chum salmon stocks that lacked nearby pink salmon survival data, we used only the R(yr − 1) forecasts as the forecast averaging model for these two stocks.

The overall MPE, the scale-independent performance measure averaged across all stocks within a species, indicated that for sockeye salmon, the naïve R(yr − 4) model performed best (had the lowest overall MPE [42%]) followed by the hybrid sibling model (66%; Figure 1). The 4-year-average model performed worst (MPE = 356%). The KF model performed better (overall MPE = 86%) than the other stock–recruitment-type models. The 4-year average plus KF forecast averaging model had an overall MPE of 221%, which

TABLE 5.—Correlations among the annual raw errors for 11 adult abundance forecasting models for 40 chum salmon stocks. See Table 4 for additional details.

	R(yr − 1)	R(yr − 4)	4-year average	Ricker	Ricker SST	Ricker AR(1)	KF	ME	Hybrid sibling	Pink salmon index
R(yr − 4)	0.37									
4-year average	0.73	0.77								
Ricker	0.39	0.73	0.72							
Ricker SST	0.38	0.71	0.73	0.96						
Ricker AR(1)	0.38	0.74	0.72	0.98	0.95					
KF	0.41	0.66	0.70	0.81	0.81	0.86				
ME	0.44	0.69	0.74	0.86	0.85	0.87	0.77			
Hybrid sibling	0.65	0.56	0.74	0.54	0.51	0.51	0.49	0.52		
Pink salmon index	0.29	0.63	0.60	0.85	0.85	0.85	0.71	0.71	0.44	
R(yr − 1) + Pink salmon index	0.76	0.63	0.82	0.80	0.79	0.79	0.72	0.73	0.67	0.84

was between the overall MPEs of the component models.

In contrast, for chum salmon, the hybrid sibling model performed best (overall MPE = 16%), followed by the naïve models (42% for the *R*[yr − 4] model, 43% for the *R*[yr − 1] model, and 47% for the 4-year-average model; Figure 1). As with sockeye salmon, the KF had the lowest overall MPE (59%) among stock–recruitment-type models, followed by the Ricker pink salmon index model (60%). The Ricker SST model had the highest overall MPE (81%). Similar to the sockeye salmon results, the overall MPE (51%) for the forecast averaging model that included the *R*(yr − 1) and Ricker pink salmon index models was intermediate between the overall MPEs of the component models. On average, across all stocks and models, overall MPEs for chum salmon were approximately half those for sockeye salmon (Figure 1).

For each performance measure, we calculated the proportion of stocks for which each model ranked first (Figures 2–5). Based on three of the four performance measures, the hybrid sibling model demonstrated the best performance for sockeye salmon stocks, but naïve models were also frequently ranked first. For chum salmon, the rankings were similar; the hybrid sibling model ranked first for a high proportion of stocks, and two naïve models also frequently ranked first (Figures 2–5).

Thus, no single forecasting model was best across all 37 sockeye salmon stocks or across all 40 chum salmon stocks; the best model depended on the particular stock and the performance measure used. Most of the 11 forecasting models ranked first for at least one stock of a given species in each of the four performance measures.

Within individual stocks, the rankings of the 11 models often depended on which performance measure

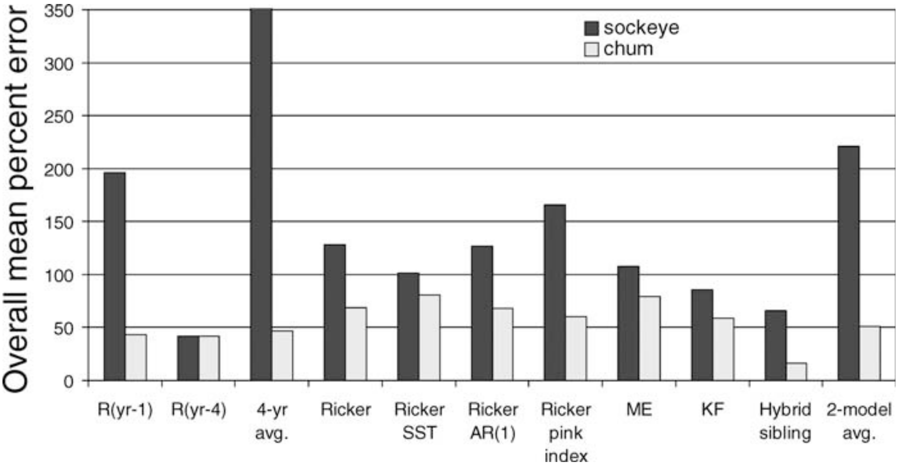

FIGURE 1.—Overall mean percent error in forecasts of the abundance of adult recruits of chum and sockeye salmon averaged across stocks and forecast years for the 11 forecasting models described in the text. Abbreviations are as follows: *R* = the return observed 1 or 4 years earlier, SST = sea surface temperature, AR[1] = first-order autoregressive, ME = mixed effects, and KF = Kalman filter.

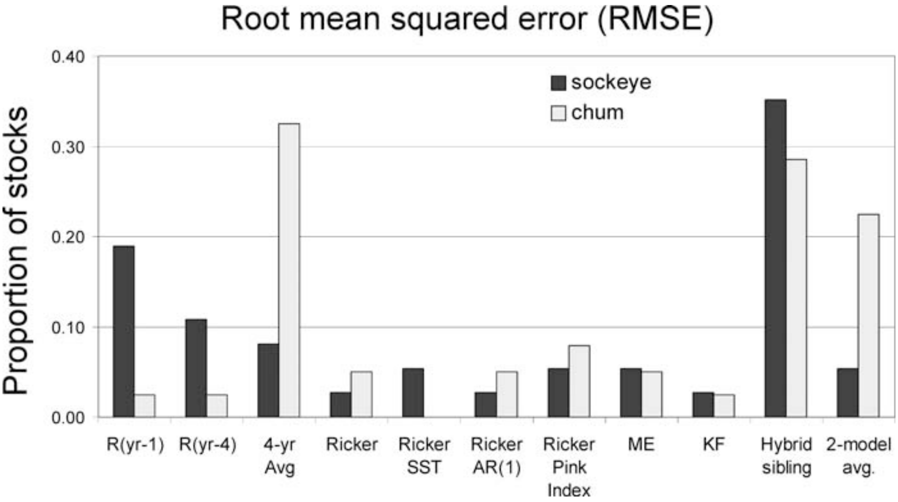

FIGURE 2.—Proportion of chum and sockeye salmon stocks for which each of the 11 forecasting models described in the text ranked first based on its root mean square error. See Figure 1 for abbreviations.

was used. Here, we provide three examples. For the Kadashan (Alaska) chum salmon stock, the ME model was ranked first based on the MPE performance measure but ranked seventh based on RMSE and sixth based on MAE (Table A.1.1). For the Weaver (BC) sockeye salmon stock, the hybrid sibling model ranked first based on RMSE but 11th based on MRE (Table A.2.1). The 4-year average model for chum salmon ranked first in 13 stocks based on RMSE, 11 stocks based on MAE, 3 stocks based on MPE, and 2 stocks based on MRE (Table A.1.1).

Although the best models varied to some extent among stocks and performance measures, some models were consistently better for some stocks. For instance, the $R(\mathrm{yr} - 1)$ model ranked first across all four performance measures for the Long Lake (BC) and Ugashik (Alaska) sockeye salmon stocks (Table A.2.1). The hybrid sibling model ranked first across all four performance measures for the Willapa Bay (Washington) and Andreafsky (Alaska) chum salmon stocks (Table A.1.1). The hybrid sibling model for sockeye salmon was best across three out of four performance measures for five stocks (Birkenhead and Weaver, BC; Cook Inlet, Naknek, and Nuyakuk,

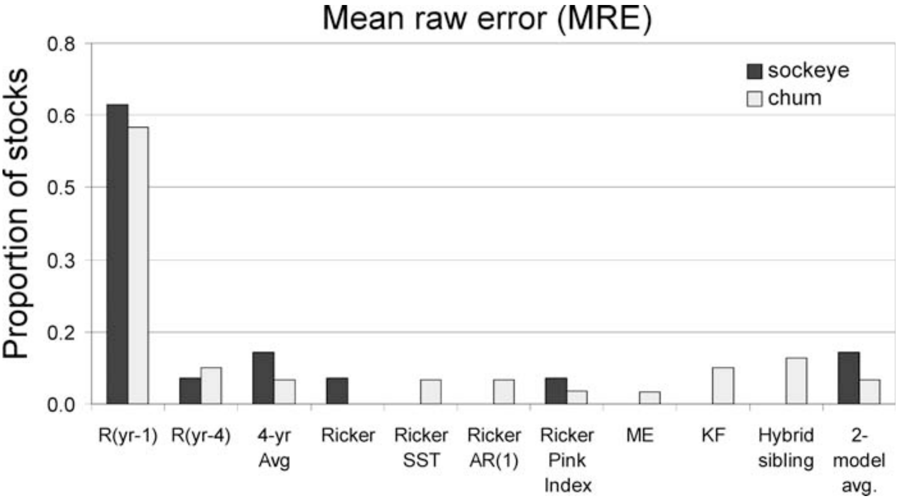

FIGURE 3.—Proportion of chum and sockeye salmon stocks for which each of the 11 forecasting models described in the text ranked first based on its mean raw error. See Figure 1 for abbreviations.

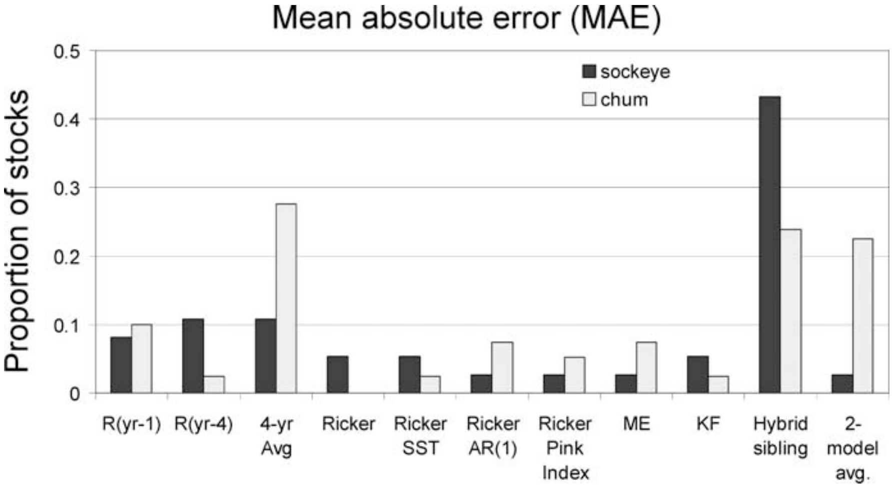

FIGURE 4.—Proportion of chum and sockeye salmon stocks for which each of the 11 forecasting models described in the text ranked first based on its mean absolute error. See Figure 1 for abbreviations.

Alaska (Table A.2.1)). The $R(\mathrm{yr}-1)$ model for chum salmon was best across all four performance measures for the Kwiniuk–Tubutulik (Alaska) stock and was best in three out of four performance measures for the Area 10 (BC) and Kamishak (Alaska) stocks (Table A.1.1).

In general, the four different performance measures reflected somewhat different characteristics of forecasting errors, as indicated by the ranking of models. Ranks based on MRE were only moderately correlated with the rankings based on the other three performance measures ($0.11 \le r \le 0.17$ for sockeye salmon; $0.31 \le r \le 0.38$ for chum salmon). The same was true for MPE ($0.11 \le r \le 0.59$ for sockeye salmon; $0.38 \le r \le 0.55$ for chum salmon), but rankings based on MAE and RMSE were quite similar ($r = 0.88$ for both species). To remove the potential effect of large outliers in forecasts, we also calculated median, rather than mean, performance measures and found that the model rankings and conclusions changed very little. To examine which models were best at avoiding large errors, we also developed a performance measure that averaged the five largest absolute errors for each stock. Ranks based on this performance measure were highly correlated with ranks based on RMSE ($r = 0.92$ for sockeye salmon; $r = 0.85$ for chum salmon). Therefore,

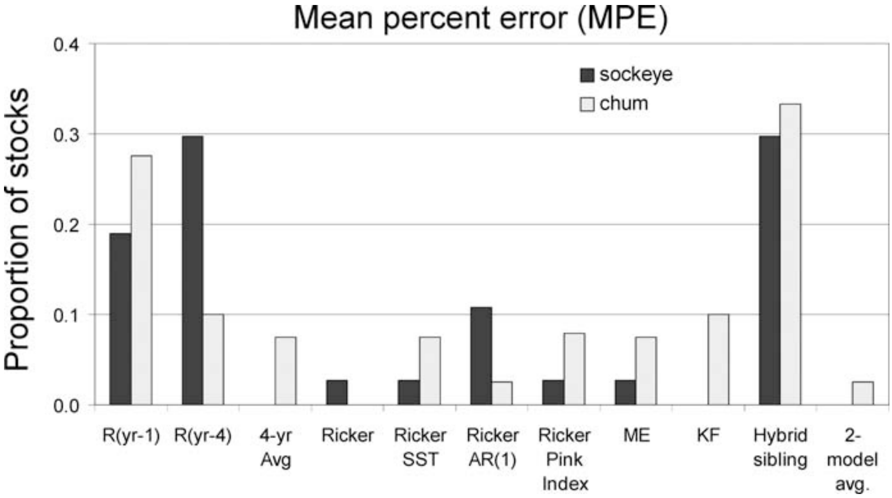

FIGURE 5.—Proportion of chum and sockeye salmon stocks for which each of the 11 forecasting models described in the text ranked first based on its mean percent error. See Figure 1 for abbreviations.

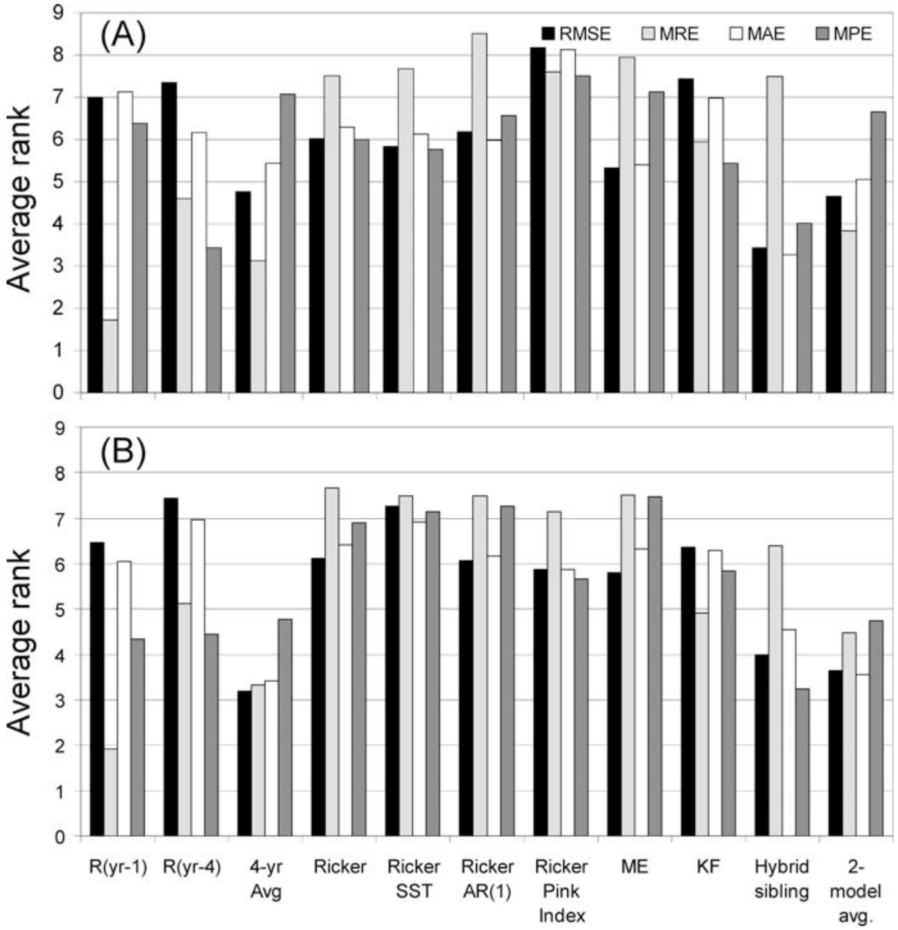

FIGURE 6.—Average ranks of the 11 forecasting models described in the text across **(A)** 37 sockeye salmon stocks and **(B)** 40 chum salmon stocks based on the results of four performance measures (root mean square error [RMSE], mean raw error [MRE], mean absolute error [MAE], and mean percent error [MPE]). An average rank of 1 is the best. See Figure 1 for abbreviations.

the ranks based on RMSE are sufficient to indicate which models avoided unusually large errors.

As with the 10 individual models, the relative performance of the forecast averaging models depended on the stock and performance measure used for ranking (Figures 2–5). The forecast averaging model for sockeye salmon (4-year average + KF) had a higher rank and a smaller forecasting error than both of its component models in 57 out of 148 cases (39%) across the four performance measures and 37 stocks. The forecast averaging model for chum salmon ($R[\text{yr} - 1] +$ Ricker pink salmon index) had a higher ranking and a smaller forecasting error than both of its component models in 46 out of 160 cases (29%) across the four performance measures and 40 stocks. Using just RMSE, the forecast averaging model was better than both of its component models in 22 of 37 sockeye salmon stocks (59%) and 22 of 40 chum salmon stocks (55%). These results suggest that some cancellation of errors and commensurate increase in accuracy can be achieved through averaging models with low or negative correlation in their raw errors.

Another measure of a forecasting model's usefulness is its average rank across all stocks within each species for a given performance measure (Figure 6). This average ranking is appropriate because Figures 2–5 indicate only the frequency with which a model ranked first, but some models ranked first for some stocks yet poorly for others (hence, mediocre on average), whereas other models consistently ranked near the top across stocks. We wished to reflect the latter. For example, the forecast averaging model for sockeye salmon performed consistently well based on RMSE, ranking second overall (Figure 6A) and at least third in

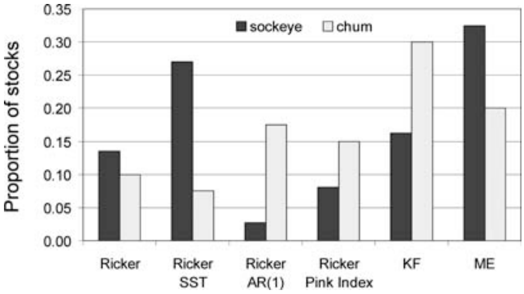

FIGURE 7.—Proportion of chum and sockeye salmon stocks for which each of the stock–recruitment-type models described in the text ranked highest among models of that type based on its root mean square error. See Figure 1 for abbreviations.

14 of 37 stocks, even though it ranked first in only a few stocks (Figures 2, 6). This high ranking was also true for the forecast averaging model for chum salmon (Figure 6B). Average ranks were generally best for the 4-year average, hybrid sibling, and forecast averaging models for both sockeye and chum salmon (Figure 6).

To characterize the relative performance of the new or recently developed stock–recruitment-type models (Ricker SST, Ricker pink salmon index, KF, and ME models), we tallied the number of times each of these models ranked first among all the stock–recruitment-type models using RMSE as the performance criterion (Figure 7). We found that these newer stock–recruitment models outperformed the standard Ricker model in 86% of the sockeye salmon stocks and 90% of the chum salmon stocks (Figure 7). Stock–recruitment models that included SST (Ricker SST and ME models) outperformed the other stock–recruitment models in 59% of the sockeye salmon stocks and 28% of chum salmon stocks. The KF and ME models performed best in about 50% of stocks for both sockeye and chum salmon. The Ricker pink salmon index model performed best in only 15% of chum salmon stocks and only 8% of sockeye salmon stocks.

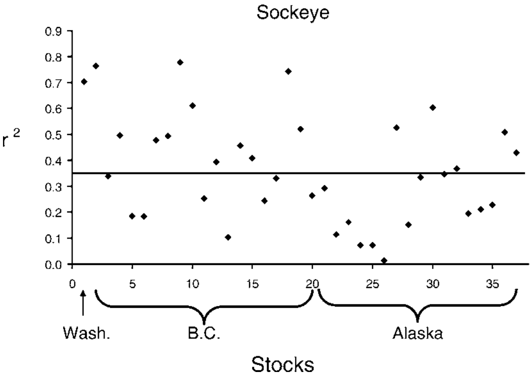

FIGURE 8.—Proportions (r^2) of the temporal variation in recruitment in 37 stocks of sockeye salmon explained by the highest-ranked forecasting model based on root mean square error. The solid horizontal line denotes the average r^2 value across all stocks. Stock numbers (x-axis) are arranged geographically from Washington State in the south (stock 1) to Bristol Bay in southwestern Alaska (stocks 29–37); see Table 1 for stock descriptions.

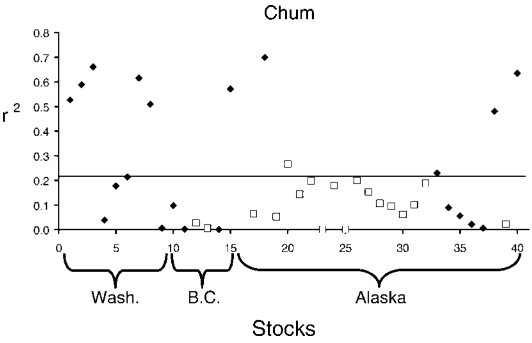

FIGURE 9.—Proportions (r^2) of the temporal variation in recruitment in 40 stocks of chum salmon explained by the highest-ranked forecasting model based on root mean square error. The diamonds pertain to stocks with age composition data, the squares to stocks with few or no age composition data. The solid horizontal line denotes the average r^2 value across all stocks. Stock numbers (x-axis) are arranged geographically from Washington State in the south (stocks 1–9) to Norton Sound in western Alaska (stocks 35–40); see Table 2 for stock descriptions.

In general, despite the wide range of forecasting model types examined, even the highest-ranked model (based on RMSE for each population) explained on average only a small proportion of the variability in recruitment (Figures 8, 9). For sockeye salmon, r^2 values for individual stocks ranged from 0.01 to 0.78 and averaged 0.36 (Figure 8). The best model typically explained a higher proportion of the variation in recruitment for individual stocks in Washington and BC (average $r^2 = 0.44$) than in Alaska (average $r^2 = 0.27$). For chum salmon, r^2 values for individual stocks ranged from 0.001 to 0.70 (average $r^2 = 0.21$; Figure 9), but no regional differences were apparent in the variation explained by the highest-ranked model based on RMSE.

One factor influencing the poor performance of preseason forecasting models for sockeye and chum salmon may be uncertainty in age-at-maturity data. More variation was explained in chum salmon stocks with age composition data (average $r^2 = 0.30$) than in chum salmon stocks without age composition data (average $r^2 = 0.11$; Figure 9). However, some of that difference may be attributed to use of the hybrid sibling model, which required age composition data and generally performed well. The hybrid sibling model ranked first in four of the eight chum salmon stocks for which r^2 equaled or exceeded 0.48 for the highest-ranking model. To remove the potential effect of alternative model types (e.g., naïve, hybrid sibling, forecast averaging) on the results, we compared average r^2 for the best-performing stock–recruitment-type model (i.e., the stock–model combinations used in Figure 7) between the 21 chum salmon stocks with age composition data and the 19 stocks with little or no age composition data. The average r^2 was 0.20 for chum salmon stocks with age composition data and 0.07 for those without age composition data.

Discussion

In our previous work (Haeseker et al. 2005) and in our present analyses, no single model emerged as consistently better at forecasting across stocks within a species for pink, chum, or sockeye salmon. This general result is consistent with past comparisons of preseason forecasting models for Pacific salmon (Fried and Yuen 1987; Noakes et al. 1990; Wood et al. 1997; Adkison 2002; Zhou 2003; Cass et al. 2006). The best model varied depending on the particular salmon stock and performance measure. Nevertheless, three general conclusions emerge from our analyses. First, regardless of performance measure, the best three models tended to include one of the naïve time series models, the forecast averaging model, and the hybrid sibling model (the last obviously only for chum and sockeye salmon). Second, even when the best of the 8 models (for pink salmon) and 11 models (for chum and sockeye salmon) was used for each separate stock, forecasting errors still tended to be large. The average proportion of variation in recruitment accounted for across all stocks within a species was only 0.36 for sockeye salmon, 0.21 for chum salmon, and 0.20 for pink salmon, which suggests that managers and harvesters should limit confidence in preseason forecasts for the foreseeable future. Third, despite their widespread usage in forecasting (often because of a lack of alternatives), stock–recruitment-type models (Ricker, Ricker AR[1], Ricker SST, KF, ME, and Ricker pink salmon index) tended to have the worst performance based on overall MPE and average ranking across performance measures. Those models also were ranked first for the smallest proportion of stocks (all $<11\%$) based on any of the four performance measures.

One background issue is important to address before we discuss our results further. Unlike most previous evaluations of preseason forecasting models for Pacific salmon (which focused on particular regions and models), the present work and our previous paper (Haeseker et al. 2005) comprehensively compared diverse models across a single, large database of 120 populations covering the three species throughout most of western North America. Altogether, we evaluated the performance of forecasting models across 1,081 stock-years for sockeye salmon, 665 stock-years for chum salmon, and 783 stock-years for pink salmon. As we synthesize results for sockeye and chum salmon below, we will compare them with our previous results for pink salmon to produce more general conclusions.

Some of these across-species differences in forecasting ability may be due to differences in life history. The amount of recruitment variation explained by the best models for chum and pink salmon (0.21 and 0.20, respectively), which enter the ocean as age-0 juveniles, was lower than that of the best model for sockeye salmon (0.36), which enter the ocean as age-1 or age-2 juveniles. However, these differences may be due to differences in other factors, such as spawning conditions, rearing habitat conditions, morphology, physiology, and quality of stock–recruitment data; intrinsic natural variability in demographic rates could also explain the differences. Comparisons in forecasting ability between these and other salmon species (e.g., coho salmon *O. kisutch* and Chinook salmon *O. tshawytscha*, which have ocean-, stream-, or lake-type life histories) could improve understanding of the influence of life history variation on recruitment uncertainty.

The long stock–recruitment time series used as data inputs in this analysis provided an opportunity to conduct a robust evaluation of forecasting model performance across 1,746 stock-years for chum and sockeye salmon. However, the methods for estimating abundance within these time series may have changed over time. For example, the bias or precision in abundance estimates over time may have been affected by changing from weir counts to aerial surveys, modifying the expansion factors used to convert abundance indices to abundance estimates, or changing the methods used to estimate stock composition within mixed-stock fisheries. However, the extent to which temporal changes in data quality may have affected forecasting model performance is unknown.

Naïve time series models performed surprisingly well for all three species despite the fact that these models did not explicitly account for biological or environmental processes (e.g., spawner abundance or ocean conditions). The $R(\mathrm{yr} - 4)$ naïve model for sockeye salmon demonstrated an overall MPE of 42%, which was 36% smaller than that of the next-best model (hybrid sibling model). All three naïve models for chum salmon demonstrated a relatively low overall

MPE (42–47%); however, the overall MPE for the hybrid sibling model was even lower (16%). For pink salmon, the *R*(yr − 2) and 2-year-average models had the lowest MPEs (82% and 103%, respectively; Haeseker et al. 2005). This general advantage of naïve models over more realistic, complex models may result from (1) the greater number of parameters requiring estimation in complex models (which creates more chances for errors in those estimates to degrade forecasts), (2) the larger number of explanatory variables used in the complex models (which similarly creates more chances for errors to corrupt forecasts; e.g., age-at-return data), (3) our lack of understanding of critical processes governing survival of salmon to recruitment (which leads to mis-specification of complex forecasting models), or (4) a combination of these.

The hybrid sibling model (Haeseker et al. 2007) was routinely in the top 3 of the 11 models for sockeye and chum salmon. This was true in terms of three performance measures: (1) overall MPE (ranked second for sockeye salmon, with an MPE = 66%, and first for chum salmon, with an MPE = 16%), (2) the proportion of stocks in which the model was the best one (ranked first or second for both sockeye and chum salmon depending on the performance measure), and (3) the average rank across stocks and performance measures (generally ranked first for sockeye salmon and first or third for chum salmon). The good performance of the hybrid sibling model is a notable departure from previous research on chum and sockeye salmon forecasting models because this new model performs better than the commonly used standard sibling model (Haeseker et al. 2007) as well as most other models (present results). One reason for the model's superior performance is that, unlike spawner–recruit models, which attempt to estimate the effects of all sources of survival rate variability between spawning and recruitment, the major sources of interannual variability may have already affected recruit abundance by the time the abundances of the earliest returns of sibling groups are estimated (Bradford 1992). Another reason for the model's success may be its structure; it uses either a naïve model or a standard sibling model, depending on the variance of residuals around the standard sibling model (details in Haeseker et al. 2007). A large residual variance triggers use of a naïve model, which we have demonstrated has surprisingly strong performance. For age-classes with strong sibling relationships, the hybrid sibling model capitalizes on the relationship's strength in forecasting the next age-class.

Several lines of evidence support the Wood et al. (1997) suggestion that a lack of data on, or large uncertainty in, age-at-return may contribute to poor performance of preseason forecasting models for chum and sockeye salmon. First, limited or absent age-at-return data preclude the use of sibling models, which demonstrated relatively good performance when applied here. Thus, less-effective models would have to be used. Second, among the chum salmon stocks, the age-based hybrid sibling model had the lowest overall MPE (16%), while the next-best case, the non-age-based *R*(yr − 4) model, had an MPE of 42%. Thus, smaller errors were produced by a forecasting model that took age structure into account. Third, for chum salmon, the amount of variation explained by the best stock–recruit model was roughly three times as great when age composition data were available as when such data were lacking. For these reasons, our analysis highlights the importance of collecting precise and accurate age composition data. Such data will reduce the problem of uncertainty in age-at-maturity data that tends to obscure patterns in brood year survival and to complicate identification and estimation of covariates for characterizing these patterns in survival. However, even complete certainty in age composition may not result in better forecasting performance. For pink salmon, stock–recruitment models generally underperformed relative to naïve models (Haeseker et al. 2005) even though there was certainty in age composition (i.e., fixed age at maturity of 2 years).

Across the three salmon species and stocks, the forecast averaging models averaged either second or third best based on RMSE, MAE, and MRE criteria (Figure 6; see also Figure 6 in Haeseker et al. 2005). While forecast averaging models demonstrated high rankings on average, they were rarely ranked first among sockeye or pink salmon stocks (Figures 2–5; see also Figures 2–5 in Haeseker et al. 2005). In contrast, the forecast averaging model for chum salmon was frequently ranked first according to the RMSE and MAE criteria.

Fried and Yuen (1987) and Noakes et al. (1990) also found that forecast averaging may improve preseason forecasts for sockeye salmon, although they evaluated a much smaller range of models and stocks. Nevertheless, salmon forecasters should choose component models carefully when deciding which to include in a forecast averaging model. Bates and Granger (1969) and Ridley (1999) provide good evidence that models with the weakest or most negatively correlated time series of forecasting errors work best in forecast averaging models to increase the precision of forecasts. We therefore suggest that forecast averaging models should not be composed solely of stock–recruitment-type models because forecasting errors of such models are strongly positively correlated for all three species

(Tables 4, 5; Table 2 of Haeseker et al. 2005). These high correlations indicate that stock–recruitment-type models are accounting for and missing the same time-varying events and that their forecasting errors will therefore not tend to cancel one another out. Instead, we suggest using correlation matrices to determine which models have low correlation among their forecast errors and which would therefore be good candidates for components of a forecast averaging model.

Although the various stock–recruitment models did not generally fare well, there were numerous individual stocks for which those models gave the best forecasts. For instance, the Ricker pink salmon index model generally performed better for chum salmon than for sockeye salmon. This result is not surprising because chum and pink salmon enter the ocean at similar times, body sizes, and ages (age-0 fry). Sockeye salmon enter the ocean as larger age-1 or age-2 smolts, and the factors that affect age-0 pink salmon may not reflect the factors that affect the larger age-1 and age-2 sockeye salmon.

The KF version of the Ricker model performed better than the standard Ricker model, according to the RMSE criterion, for 41% of sockeye salmon stocks and 53% of chum salmon stocks. This is consistent with expectations based on the work of Peterman et al. (2000), who demonstrated through simulations that the KF model is relatively effective at tracking changing productivity regimes. The enhanced performance of the KF model may reflect substantial changes in underlying productivity over time in sockeye and chum salmon stocks (as opposed to year-to-year variation). We saw evidence of underlying trends in the Ricker *a* parameter for both sockeye and chum salmon (not shown here).

The multistock ME model also performed better than other stock–recruitment models. According to the RMSE criterion, the ME model outperformed the Ricker SST model in 57% of sockeye salmon stocks, 65% of chum salmon stocks, and 53% of pink salmon stocks. Both models incorporate SST, but during parameter estimation the ME model simultaneously uses information from all stocks of a given species. These results are consistent with those of Mueter et al. (2002a) and Su et al. (2004), indicating that the multistock approach can provide more precise and accurate estimates of model parameters, translating into better forecasting accuracy, as demonstrated here and in our previous analysis (Haeseker et al. 2005).

In conclusion, our comprehensive evaluations of a wide range of preseason forecasting models across 120 populations of pink, chum, and sockeye salmon show that no single model is universally appropriate for a given species. However, the hybrid sibling model for chum salmon provides by far the lowest average percent forecasting error (16%) of any model for any of these three species. Another category of consistently highly ranked models is the naïve models, which forecast recruitment based on recruitment 1, 2, or 4 years before or 2- or 4-year moving averages. Nonetheless, the proportion of annual variation in recruitment that is accounted for by the best stock-specific forecasting models is generally quite small (<36%). This figure might increase for particular populations through future research, but a large improvement in forecasting ability seems unlikely given the breadth of stocks and models examined to date. Even applying a correction for prediction bias (Chen 2004) only improves forecasts by a few percentage points, and we have demonstrated that more complex models are not necessarily superior. Therefore, better in-season updates and adjustments to fishing regulations and a cautious approach to the opening and closing of fisheries will remain high priorities.

Acknowledgments

We thank Brian Pyper, Mike Bradford, and the late Dan Ware for helpful discussions and critique, and we thank Brigitte Dorner for comments on a draft manuscript. Thanks also to the numerous biologists and technicians in various management agencies (Tables 1, 2) for gathering and processing the lengthy time series data and providing them to us. The COADS temperature data were provided by the Climate Diagnostics Center (Cooperative Institute for Research in Environmental Sciences, National Oceanic and Atmospheric Administration, Boulder, Colorado), from their web site. This research was supported by a Strategic Projects Grant from the Natural Sciences and Engineering Research Council of Canada to R.M.P., C.C.W., Mike Bradford, and Dan Ware.

References

Adkison, M. D. 2002. Preseason forecasts of pink salmon harvests in Southeast Alaska using Bayesian model averaging. Alaska Fishery Research Bulletin 9:1–8.

Adkison, M. D., and R. M. Peterman. 2000. Predictability of Bristol Bay, Alaska, sockeye salmon returns one to four years in the future. North American Journal of Fisheries Management 20:69–80.

Adkison, M. D., R. M. Peterman, M. F. Lapointe, D. M. Gillis, and J. Korman. 1996. Alternative models of climatic effects on sockeye salmon, *Oncorhynchus nerka*, productivity in Bristol Bay, Alaska, and the Fraser River, British Columbia. Fisheries Oceanography 5:137–152.

Alaska Department of Fish and Game. 1981. Preliminary forecasts and projections for 1981 Alaskan salmon fisheries. Alaska Department of Fish and Game, Information Leaflet 190, Anchorage.

Alaska Department of Fish and Game. 2006. 2006 Prince William Sound and Copper/Bering River area salmon fishery forecasts. Available: www.cf.adfg.state.ak.us/region2/finfish/salmon/pws/pwsfor06.php. (November 2006).

Anderson, A. D., and D. D. Bailey. 1974. The return of chum salmon stocks to the Johnstone Strait–Fraser River study area, and prospects for 1974. Environment Canada, Technical Report Series PAC/T-14-15, Vancouver.

Bates, J. M., and C. W. J. Granger. 1969. The combination of forecasts. Operational Research Quarterly 20:451–468.

Beauchamp, J. J., and J. S. Olson. 1973. Correction for bias in regression estimates after logarithmic transformations. Ecology 54:1403–1407.

Bigler, B. 1985. Kotzebue Sound chum salmon (*Oncorhynchus keta*) escapement and return data. Alaska Department of Fish and Game, Technical Data Report 149, Kotzebue.

Bocking, R. C., and R. M. Peterman. 1988. Preseason forecasts of sockeye salmon (*Oncorhynchus nerka*): comparison of methods and economic considerations. Canadian Journal of Fisheries and Aquatic Sciences 45:1346–1354.

Bradford, M. J. 1992. Precision of recruitment predictions from early life stages of marine fishes. U.S. National Marine Fisheries Service Fishery Bulletin 90:439–453.

Campbell, R. D., A. R. Shaul, M. R. Witteveen, and J. J. Dinnocenzo. 1999. South Peninsula annual management report, 1998. Alaska Department of Fish and Game, Regulatory Information Report 4K99-29, Kodiak.

Cass, A., M. Folkes, C. Parken, and C. C. Wood. 2006. Preseason run size forecasts for Fraser River sockeye for 2006. Canadian Science Advisory Secretariat, Research Document 2006/60, Nanaimo, British Columbia. Available: www.dfo-mpo.gc.ca/csas. (February 2007).

Chatfield, C. 2000. Time series forecasting. Chapman and Hall, New York.

Chen, D. G. 2004. Bias and bias correction in fish recruitment prediction. North American Journal of Fisheries Management 24:724–730.

Clark, J. H. 2001a. Biological escapement goal for Andreafsky River chum salmon. Alaska Department of Fish and Game, Regional Information Report 3A01-07, Anchorage.

Clark, J. H. 2001b. Biological escapement goals for Kwiniuk and Tubutulik chum salmon. Alaska Department of Fish and Game, Regional Information Report 3A01-08, Anchorage.

Clark, J. H. 2001c. Biological escapement goal for chum salmon in District 1 of Norton Sound. Alaska Department of Fish and Game, Regional Information Report 3A01-09, Anchorage.

Clark, J. H., and G. H. Sandone. 2001. Biological escapement goal for Anvik River chum salmon. Alaska Department of Fish and Game, Regional Information Report 3A01-06, Anchorage.

Clemen, R. T. 1989. Combining forecasts: a review and annotated bibliography. International Journal of Forecasting 5:559–583.

Cross, B. A., and D. C. Gray. 1999. A synopsis and critique of forecasts of sockeye salmon returning to Bristol Bay in 1998. Alaska Department of Fish and Game, Regional Information Report 2A99-03, Anchorage.

Eggers, D. M. 2001. Biological escapement goals for Yukon River fall chum salmon. Alaska Department of Fish and Game, Regional Information Report 3A01-10, Anchorage.

Fried, S. M., and H. J. Yuen. 1987. Forecasting sockeye salmon (*Oncorhynchus nerka*) returns to Bristol Bay, Alaska: a review and critique of methods. Canadian Special Publication of Fisheries and Aquatic Sciences 96:273–279.

Groot, C., and L. Margolis. 1991. Pacific salmon life histories. University of British Columbia Press, Vancouver.

Haeseker, S. L., B. Dorner, R. M. Peterman, and Z. Su. 2007. An improved sibling model for forecasting chum salmon and sockeye salmon abundance. North American Journal of Fisheries Management 27:634–642.

Haeseker, S. L., R. M. Peterman, Z. Su, and C. C. Wood. 2005. Retrospective evaluation of preseason forecasting models for pink salmon. North American Journal of Fisheries Management 25:897–918.

Holt, C. A., and R. M. Peterman. 2006. Missing the target: uncertainties in achieving management goals in fisheries on Fraser River, British Columbia, sockeye salmon (*Oncorhynchus nerka*). Canadian Journal of Fisheries and Aquatic Sciences 63:2722–2733.

Mueter, F. J., R. M. Peterman, and B. J. Pyper. 2002a. Opposite effects of ocean temperature on survival rates of 120 stocks of Pacific salmon (*Oncorhynchus* spp.) in northern and southern areas. Canadian Journal of Fisheries and Aquatic Sciences 59:456–463.

Mueter, F. J., D. M. Ware, and R. M. Peterman. 2002b. Spatial correlation patterns in coastal environmental variables and survival rates of salmon in the northeast Pacific Ocean. Fisheries Oceanography 11:205–218.

Murphy, R. L., A. R. Shaul, and J. J. Dinnocenzo. 1999. North Alaska Peninsula commercial salmon annual management report, 1998. Alaska Department of Fish and Game, Regulatory Information Report 4K99-34, Kodiak.

Noakes, D. J., D. W. Welch, M. Henderson, and E. Mansfield. 1990. A comparison of preseason forecasting methods for returns of two British Columbia sockeye salmon stocks. North American Journal of Fisheries Management 10:46–57.

Owen, D. L., and D. R. Sarafin. 1999. Chignik management area annual finfish management report, 1996. Alaska Department of Fish and Game, Regulatory Information Report 4K99-33, Kodiak.

Peterman, R. M. 1981. Form of random variation in salmon smolt-to-adult relations and its influence on production estimates. Canadian Journal of Fisheries and Aquatic Sciences 38:1113–1119.

Peterman, R. M. 1982. Model of salmon age structure and its use in preseason forecasting and studies of marine survival. Canadian Journal of Fisheries and Aquatic Sciences 39:1444–1452.

Peterman, R. M. 1987. Review of the components of recruitment of Pacific salmon. Pages 417–429 *in* M. J. Dadswell, R. J. Klauda, C. M. Moffitt, R. L. Saunders, R. A. Rulifson, and J. E. Cooper, editors. Common strategies of anadromous and catadromous fishes.

American Fisheries Society, Symposium 1, Bethesda, Maryland.

Peterman, R. M., B. J. Pyper, and J. A. Grout. 2000. Comparison of parameter estimation methods for detecting climate-induced changes in productivity of Pacific salmon (*Oncorhynchus* spp.). Canadian Journal of Fisheries and Aquatic Sciences 57:181–191.

Peterman, R. M., B. J. Pyper, M. F. Lapointe, M. D. Adkison, and C. J. Walters. 1998. Patterns of covariation in survival rates of British Columbian and Alaskan sockeye salmon (*Oncorhynchus nerka*) stocks. Canadian Journal of Fisheries and Aquatic Sciences 55:2503–2517.

Pyper, B. J., F. J. Mueter, and R. M. Peterman. 2005. Across-species comparisons of spatial scales of environmental effects on survival rates of Northeast Pacific salmon. Transactions of the American Fisheries Society 134:86–104.

Pyper, B. J., F. J. Mueter, R. M. Peterman, D. J. Blackbourn, and C. C. Wood. 2001. Spatial covariation in survival rates of Northeast Pacific pink salmon (*Oncorhynchus gorbuscha*). Canadian Journal of Fisheries and Aquatic Sciences 58:1501–1515.

Pyper, B. J., F. J. Mueter, R. M. Peterman, D. J. Blackbourn, and C. C. Wood. 2002. Spatial covariation in survival rates of Northeast Pacific chum salmon. Transactions of the American Fisheries Society 131:343–363.

Rice, J. A. 1995. Mathematical statistics and data analysis. Duxbury Press, Belmont, California.

Ridley, D. 1999. Optimal antithetic weights for lognormal time series forecasting. Computers and Operations Research 26:189–209.

Ryall, P., C. Murray, V. Palermo, D. Bailey, and D. Chen. 1999. Status of Clockwork chum salmon stock and review of the Clockwork management strategy. Canada Department of Fisheries and Oceans, Research Document 99/169, Ottawa.

Shao, J. 1993. Linear model selection by cross validation. Journal of the American Statistical Association 88:486–494.

Su, Z., R. M. Peterman, and S. L. Haeseker. 2004. Spatial hierarchical Bayesian models for stock–recruitment analysis of pink salmon (*Oncorhynchus gorbuscha*). Canadian Journal of Fisheries and Aquatic Sciences 61:2471–2486.

Washington Department of Fish and Wildlife. 2006. Management of Puget Sound chum salmon. Available: wdfw.wa.gov/fish/chum/chum-5c.htm. (November 2006).

Wood, C. C., D. T. Rutherford, D. Peacock, S. Cox-Rogers, and L. Jantz. 1997. Assessment of recruitment forecasting methods for major sockeye and pink salmon stocks in northern British Columbia. Canadian Technical Report of Fisheries and Aquatic Sciences 2187.

Zhou, S. 2003. Application of artificial neural networks for forecasting salmon escapement. North American Journal of Fisheries Management 23:48–59.

Appendixes: Recruitment Models for Chum and Sockeye Salmon Stocks

(Available online at afs.allenpress.com. See links in Abstract file.)