North American Journal of Fisheries Management 29:423–433, 2009

DOI: 10.1577/M07-055.1

[Article]

The Use of Generalized Additive Models for Forecasting the Abundance of Queets River Coho Salmon

SHIZHEN WANG*[1] AND GARY MORISHIMA
Quinault Management Center, 3010 77th Avenue SE, Suite 104, Mercer Island, Washington 98040, USA

RISHI SHARMA
*Columbia River Inter-Tribal Fish Commission,
729 Northeast Oregon Street, Suite 200, Portland, Oregon 97232, USA*

LARRY GILBERTSON
*Quinault Fisheries Department, Quinault Indian Nation,
Post Office Box 189, Taholah, Washington 98587, USA*

Abstract.—We examined three types of models for preseason forecasting of the abundance of Queets River coho salmon *Oncorhynchus kisutch*: (1) a simple model in which estimates of smolt production are multiplied by projected marine survival rates, (2) a Ricker spawner–recruitment model, and (3) a regression model relating log-transformed adult recruitment to smolt production. Each type of model was formulated with and without environmental variables that influence production and survival. We attempted to use a nonparametric generalized additive model (GAM) to guide the selection of the environmental variables and the form of the regression model. The GAM model was derived through a stepwise selection strategy based on the Akaike information criterion. Parametric approximate models were developed for each selected GAM model, and their performance was compared with postseason estimates of abundance using three criteria: the mean absolute percentage error, the largest absolute percentage error, and the probability of being included in the 90% prediction interval. This paper shows that the GAM approach is useful in constructing forecasting models by identifying promising relationships with predictor variables and improving abundance forecasts through the incorporation of environmental variables.

Pacific salmon *Oncorhynchus* spp. have long supported commercial, recreational, and native fisheries along the eastern Pacific coast from California to Alaska. Pacific salmon are semelparous, anadromous species with a high degree of fidelity to their natal streams. Semelparity means that individual fish have only one opportunity to spawn. The high degree of homing fidelity has led to the development of individual spawning populations, termed stocks, that have become adapted to local environmental conditions.

Conservation of these species is directed at constraining exploitation to the levels required to sustain and conserve individual stocks. In the Pacific Northwest, stock-based management is also required to comply with obligations stemming from treaties between the United States and several Indian tribes, the Pacific Salmon Treaty between the United States and Canada (PSC 2000), and the responsibility to protect fish listed under the U.S. Endangered Species Act.

For coho salmon *O. kisutch* originating in the rivers of southern British Columbia, Washington, and Oregon, the fishery planning process in both domestic and international forums is driven by annual stock abundance forecasts. Since much of the harvest is taken by mixed-stock fisheries, catch levels are constrained by the natural stocks that can sustain the lowest level of exploitation. The inability of mixed-stock fisheries to access more abundant stocks that can withstand higher rates of exploitation results in the so-called "mixed-stock problem." Although most evident in marine fisheries, the mixed-stock problem also occurs in river fisheries owing to overlaps in run timing between natural and hatchery coho salmon and other species. The mixed-stock problem makes it particularly important for the abundance forecasts of limiting stocks to be as accurate as possible.

Coho salmon from the Queets River, Washington (Figure 1), have a three-year life history pattern. Free-swimming fry emerge from eggs in spring and reside in the stream until they migrate to the sea as smolts in approximately May–June of the following year. After a year and a half at sea, the fish mature and return to

* Corresponding author: swang@dfg.ca.gov
[1] Present address: Pacific States Marine Fisheries Commission, 4665 Lampson Avenue, Suite C, Los Alamitos, California 90720, USA.

Received March 20, 2007; accepted August 7, 2008
Published online April 20, 2009

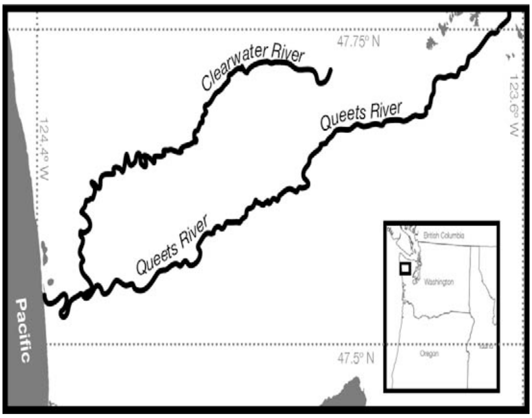

FIGURE 1.—Map of the Queets River watershed.

spawn in November and December or late October to early January.

Queets River coho salmon have limited ocean fisheries off the Washington coast since the 1970s (PFMC 2006) and are identified as a key management unit under the 2002 Southern Coho Agreement reached by the Pacific Salmon Commission.[2]

Because of the significance of this stock, estimates of smolt production, adult spawning escapements (parent spawners), and marine survivals have been collected for Queets River coho salmon since the late 1970s. For several years, the annual preseason abundance forecasts of the number of natural adults were based on estimates of smolt production and recent average marine survival rates. The errors in these forecasts were so large that investigations were undertaken to improve them by incorporating information on environmental factors that appeared to be related to survival.

The annual abundance of coho salmon often varies substantially because the progeny produced by spawning adults are subjected to unstable conditions in rivers, estuaries, and the ocean. The impacts of environmental conditions on coho salmon production and survival have been demonstrated by many studies. Lawson et al. (2004) found that annual air temperatures and second-year winter flows were related to smolt production among coastal coho salmon in Washington and Oregon. Nickelson et al. (1992) found that side channel habitat was critical to the survival of coastal coho salmon in Oregon, providing shelter from the high-flow events that occur regularly in coastal streams. Ocean conditions, such as sea surface temperatures and upwelling during the time of ocean entry, have been found to affect the marine survival rate of coho salmon (e.g., Nickelson 1986; Pearcy 1992; Coronado and Hilborn 1998; Ryding and Skalski 1999; Cole 2000; Koslow et al. 2002; Logerwell et al. 2003). Over much of the Pacific Northwest, the abundance and survival of Pacific salmon have also been shown to fluctuate on a decadal or interdecadal scale (Francis and Hare 1994; Hare and Francis 1995; Mantua et al. 1997; Hare et al. 1999).

In this paper, we investigate the ability of generalized additive models (GAMs) to improve abundance forecasts for naturally produced coho salmon from the Queets River. The GAM is a flexible and powerful tool that can detect relationships between the response variable and the predictor variables (Hastie and Tibshirani 1990).

Methods

Alternative Forecasting Models

We evaluated three basic types of abundance forecasting models: (1) a simple model in which estimates of smolt production are multiplied by projections of marine survival rates, (2) a Ricker spawner–recruitment model, and (3) a regression model that relates log-transformed adult production to smolt production. We began with the basic forms of these models, then modified them by incorporating environmental variables.

Model 1: marine survival.—In the first type of model, abundance is forecast as the product of a smolt production estimate and a projected marine survival rate, that is,

$$R_t = r_t \cdot \mathrm{Smolts}_t, \qquad (1A)$$

where R_t, r_t, and Smolts_t are adult abundance, the marine survival rate, and the smolt production estimate, respectively, for brood year t. Historically, r_t was estimated by a recent 5-year moving average of marine survival rates.

Environmental variables are explicitly incorporated into this type of model by reformulating it as

$$R_t = r'_t \cdot \mathrm{Smolts}_t, \qquad (1B)$$

in which r'_t, is predicted by a regression model with environmental variables, that is,

$$r'_t = \alpha + \sum_{j=1}^{k} f_j(E_{t,j}) + \varepsilon_t,$$

[2] Prior to the late 1990s, Canadian fisheries off the west coast of Vancouver Island exerted substantial pressure on Queets River coho salmon. Since that time, the pressure has been reduced by severe restrictions aimed at conserving interior Fraser River coho salmon. Consequently, Queets River coho salmon have not been a critical limiting stock for Pacific Fishery Management Council fishery management in recent years.

where α is a constant, f is a nonparametric smoothing function, $E_{t,j}$ is the observed value of the jth ocean environmental variable (such as sea surface temperature, sea level pressure, or an upwelling index) that affects the survival of the smolts from brood year t, and ε_t a random error term with mean 0 and variance σ^2.

Model 2: Ricker stock–recruitment.—The Ricker model (Ricker 1975) is defined by two parameters, α and β. Usually, these parameters are estimated from a data set comprising the spawning population (S_t) and subsequent recruitment to adult using linear regression, that is,

$$\log_e(R_t/S_t) = \log_e\alpha - \beta S_t + \varepsilon_t. \qquad (2A)$$

The Ricker model incorporating environmental covariates has the form

$$\log_e(R_t/S_t) = \alpha + f_0(S_t) + \sum_{j=1}^{k} f_j(E_{t,j}) + \varepsilon_t, \qquad (2B)$$

where f is a nonparametric smoothing function and $E_{t,j}$ is the observed value of the jth freshwater or ocean environmental variable (such as annual air temperature, second-winter water flow, sea surface temperature, sea level pressure, or an upwelling index) for brood year t.

Model 3: smolt to adult.—The third type of model has the basic form

$$\log_e(R_t) = \alpha + f_0(\mathrm{Smolts}_t) + \varepsilon_t. \qquad (3A)$$

With environmental covariates, this model has the form

$$\log_e(R_t) = \alpha + f_0(\mathrm{Smolts}_t) + \sum_{j=1}^{k} f_j(E_{t,j}) + \varepsilon_t, \qquad (3B)$$

where $E_{t,j}$ is as before.

GAM and Model Selection Methods

Instead of building different types of regression model (e.g., linear, exponential, polynomial, etc.) and choosing the best-fitting one, we employed GAM models to explore the relationships between the response variable and various predictor variables for models 1B, 2B, 3A, and 3B. The GAM algorithm is based on a general unspecific (nonparametric) function f that relates the values of the predicted (transformed) response variable (y) to the values of the predictor variable (x). We estimated f by a cubic smoothing spline $\hat{f}$ that minimizes the penalized residual sum of squares

$$\sum [y_i - \hat{f}(x_i)]^2 + \lambda \int [\hat{f}''(x)]^2 dx,$$

where λ is a smoothing parameter (Hastie and Tibshirani 1990). When the smoothing spline is linear, the integral $[\hat{f}''(x)]^2\,dx$ becomes 0, resulting in a least-squares-based general linear model.

The environmental variables in models 1B, 2B, and 3B were selected via a stepwise selection strategy using the full data set presented in Table 1. The selection procedures were carried out with Splus software (Insightful, Inc.; Venables and Ripley 1999). Each environmental variable could be dropped, entered linearly, or entered through a smoothing spline with 2, 3, or 4 degrees of freedom. The number of degrees of freedom of a smoothing spline is related to the number of times the function changes direction. Because the data set available for the analysis was relatively small, we chose not to incorporate interaction terms into the GAM models to avoid the risk of overfitting.

The Akaike information criterion (AIC; Akaike 1974) was employed to compare the performance of the alternative forms of each model. The AIC is calculated as $-2 \cdot \log$ likelihood $+ 2p$, where p is the number of parameters in the fitted model; a lower AIC value often implies a better fit.

Generalized additive model techniques provide a powerful means of exploring the relationships between the response variable and the predictor variables. However, a parametric model is preferred to a nonparametric one for the purpose of prediction because of the difficulties in applying and interpreting the latter. We fit a parametric regression model to each selected GAM model based on the functional forms suggested by the nonparametric fit. For example, we chose a parametric polynomial model if the selected GAM model suggested a quadratic or cubic transformation for the predictor variable. The fit of the parametric regression model to the selected GAM model was evaluated by means of a chi-square test on the difference of their deviances (residual sums of squares). The deviance for each parametric model was obtained by fitting it with the same full data set as for the selected GAM model using the least-squares method.

Validation of the Parametric Models

The predictive performance of each parametric regression model was evaluated through both leave-one-out cross validations (Efron and Tibshirani 1993) and hindcasting validations. In the leave-one-out method, the model was fit with a data set excluding the data for the year to be forecast. The resulting model was then used to forecast adult abundance for that year. This method utilizes the maximum amount of historical data in each validation process to improve estimation

TABLE 1.—Data for Queets River coho salmon abundance forecasting. Abbreviations are as follows: AnnTemp = the average monthly mean air temperature (°C) for the 16-month period from December of the brood year to March of the smolt year; WinFlow = the average monthly mean stream flow (cf^3/s) from November prior to the smolt year to March of the smolt year; SSTAMJ = the average monthly mean sea surface temperature (°C) from April to June of the smolt year; UpwellAMJ = the average monthly mean upwelling index from April to June of the smolt year; and SLPAMJ = the average monthly mean sea level pressure (millibars) from April to June of the smolt year.

Brood year	Adult abundance	Parent spawners	Smolt production	AnnTemp	WinFlow	SSTAMJ	UpwellAMJ	SLPAMJ
1979	19,403	6,800	168,300	9.35	8,331	11.53	4.00	1,017.36
1980	9,302	4,700	135,500	9.17	8,674	10.88	44.33	1,016.99
1981	21,134	4,800	324,272	8.76	9,027	12.43	24.67	1,016.80
1982	13,141	7,000	243,031	9.06	8,651	11.14	6.00	1,017.20
1983	13,121	2,282	153,741	8.05	4,583	10.97	25.33	1,018.22
1984	14,134	8,900	266,935	8.07	6,232	11.61	11.33	1,017.01
1985	7,328	3,839	120,650	8.76	6,868	11.54	17.67	1,017.72
1986	10,403	4,850	195,795	9.08	5,049	11.66	2.33	1,016.17
1987	14,797	4,531	258,711	8.37	6,248	11.61	21.67	1,017.81
1988	25,884	4,217	375,977	8.51	7,846	11.83	11.00	1,016.66
1989	11,370	3,871	190,703	8.79	10,412	11.10	20.00	1,018.23
1990	7,387	5,163	252,158	8.97	7,210	12.98	21.33	1,016.45
1991	1,689	6,023	146,315	9.26	4,494	12.44	−11.33	1,014.46
1992	9,853	5,907	243,826	8.59	6,380	12.25	5.67	1,016.53
1993	11,013	4,876	185,600	8.76	9,328	12.18	12.00	1,016.03
1994	1,797	1,039	98,742	9.06	9,599	11.72	−9.67	1,016.97
1995	4,957	5,626	339,787	8.72	9,410	12.23	−1.00	1,016.42
1996	5,244	8,784	144,869	9.27	5,758	12.58	26.00	1,016.51
1997	8,781	1,450	76,077	8.99	12,588	10.79	12.00	1,018.75
1998	29,359	3,996	322,395	8.19	7,669	11.72	7.00	1,017.52
1999	16,132	4,687	256,919	8.35	3,662	11.28	13.00	1,018.53
2000	13,337	7,939	397,716	8.00	9,245	10.91	12.33	1,017.87
2001	10,234	23,495	372,075	8.57	7,592	12.09	10.33	1,015.49
2002	9,733	13,833	384,103	9.19	6,389	12.61	11.00	1,017.40

of the model parameters, an obvious advantage when working with a short data series.

In the hindcasting method, each predictive model was parameterized using only the data that would have been available at the time the forecast would have been made (e.g., we predicted the adult abundance for 1993 with data from 1979 to 1992, that for 1994 with the data from 1979 to 1993, and so on). The results from this method would be expected to be more representative of its performance in real-world applications.

We used nonparametric bootstrapping (Efron and Tibshirani 1993) to estimate a 90% prediction interval (PI) for each adult abundance forecast. Bootstrapping for model 1A involved the following steps: (1) resampling (with replacement) the marine survival rates of the 5 years immediately prior to the forecast year, (2) calculating the mean marine survival rate from the sample, (3) computing adult abundance by multiplying the corresponding smolt production estimate by the mean marine survival rate obtained from step 2, (4) repeating steps 1–3 1,000 times, and (5) calculating the 90% PI for the adult abundance forecast based on the values obtained from steps 1–4. For all other models, bootstrapping involved the following procedures: (1) resampling (with replacement) the residuals from the model fit, (2) forming new values of the response variable by adding the sampled residuals to the values generated by the forecasting model, (3) calculating regression coefficients by refitting the model with the new values of the response variable and the original values of the predictor variable(s), (4) computing adult abundance using the model parameterized with the regression coefficients obtained from step 3, (5) repeating steps 1–4 1,000 times, and (6) calculating the 90% PI of the adult abundance forecast based on the values obtained from steps 1–5.

Methods for Fitting the Parametric Models

We estimated the parameters for each parametric approximate model by the method of least squares (Neter et al. 1996). The statistical independence of the error terms across years was examined based on the model residuals by the Durbin–Watson test or plots of the correlation and partial correlation functions when the Durbin–Watson test was indeterminate. Autocorrelation of the error terms can bias the recruitment estimates from the stock–recruitment model and hence requires corrective measures (Walters 1985, 1990; Hilborn and Walters 1992; Quinn and Deriso 1999). However, we did not find statistically significant correlations between the error terms for the models fit with the full data set (during model construction) or those fit with partial data sets (during the validations).

Because we used bootstrapping to calculate prediction intervals for the abundance forecasts, we chose not to report the diagnostic results for the normality and heteroscedasticity (nonconstancy of variance) of the error terms of the parametric regression models. Nonnormality and heteroscedasticity of the error terms can affect the accuracy of statistical inferences about model parameter estimates (e.g., standard errors and hypothesis tests) but not the accuracy of the least-squares estimates of the parameters of a linear model based on the normal theory (Neter et al. 1996).

Bias Correction for the Predicted Abundance

Additive models with log-transformed response variables (e.g., models 2A, 2B, 3A, and 3B) are appropriate for a combination of independent mortality processes. Log-transforming the response variable converts multiplicative mortality processes to a series of additive mortality processes and often helps stabilize the variance of the error terms of the model. However, log-transformation of the response variable can introduce bias into the estimates of the untransformed response variable if these estimates are obtained by merely taking antilogarithms of the fitted values (Bradu and Mundlak 1970; Duan 1983). We corrected the bias of the adult abundance forecasts caused by log transformation using the method-of-smearing estimate (Duan 1983), that is, by multiplying the adult abundance forecasts obtained by taking the antilogarithms of the fitted values by

$$\left[\sum_{i=1}^{n}\exp(\hat{\varepsilon}_i)\right]/n,$$

where n is the number of observations and $\hat{\varepsilon}_i$ are the estimated residuals.

Comparing the Performance of the Alternative Models

The abundance forecasting models were compared in terms of their postseason estimates by means of three statistics: (1) the mean absolute percentage error, (2) the largest annual absolute percentage error among all predicted years, and (3) the 90% PI coverage probability.

The absolute percentage error is defined as

$$\frac{|\hat{R}_t - R_t|}{R_t} \times 100,$$

where $\hat{R}_t$ is the predicted adult abundance for year t and R_t is the postseason estimate of adult abundance for year t. The mean absolute percentage error (MAPE) is defined as the average of the absolute percentage errors over the number of prediction years. The average deviation (regardless of sign) is reflected by the magnitude of the MAPE statistic; the smaller the MAPE, the closer the model predictions are to the estimates of actual abundance.

The largest annual absolute percentage error (LAPE) represents the largest deviation of the predicted values from the postseasonally observed abundance among all prediction years. The LAPE is an important criterion for evaluating the performance of forecasting models for Pacific salmon because the consequences of errors in a forecast are likely to be more serious as the magnitude of error increases. For instance, overfishing that results from the overestimation of abundance can cause low spawning escapement and reduce future production.

The 90% PI coverage probability is the proportion of years for which the observed adult abundance falls within the 90% PI for the forecast. The greater the coverage probability, the more reliable the model.

A rank was assigned to each of the six models for each performance metric. If two or more metric values were tied, we assigned the average rank to each tied value. The ranks were then summed for each model; the best model was the one with the lowest rank-sum.

Data Sources

The Quinault Department of Fisheries has been studying Queets River coho salmon for over three decades. A data series consisting of annual estimates of spawning escapement (parent spawners), smolt production, total adult abundance, and marine survival rates has been compiled for the 1979–2002 brood years.

Escapement was estimated from redd counts and other data collected during spawning ground surveys throughout the watershed. Smolt production was estimated through mark–recapture experiments. Total adult abundance was estimated by expanding escapements for terminal and ocean harvest based on exploitation rates derived from cohort analyses of coded-wire-tagged fish. Marine survival rates were calculated by dividing adult abundance by the corresponding smolt production estimates.

Based on a review of the literature on the relationships of salmon production and survival to environmental factors, the following environmental variables were used in our analyses (the full data set is presented in Table 1):

Annual air temperature (AnnTemp) was calculated as the arithmetic average of monthly mean air temperatures over the 16-month period from December of the brood year to March of the smolt year. Monthly mean air temperatures were obtained from the Western Regional Climate Center (http://www.wrcc.dri.edu) for

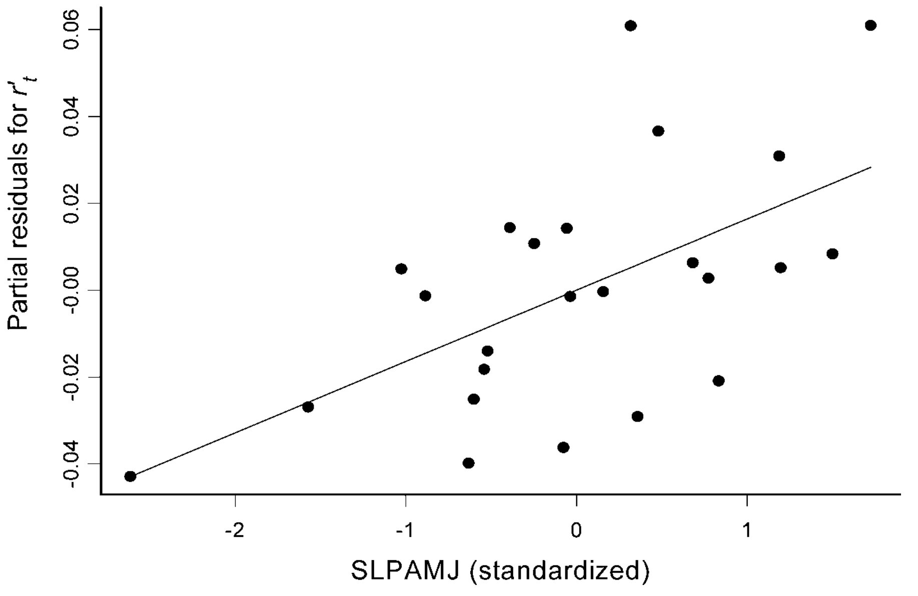

FIGURE 2.—Fit of the GAM model selected for model 1B. Partial residuals are the residuals that result from removing the effect of the predictor variable, in this case SLPAMJ (sea level pressure from April to June during the smolt year of Queets River coho salmon); SLPAMJ was standardized by subtracting the mean value of all observations and dividing the result by the standard deviation. The solid line represents the component smoothing function, which is linear for this model.

the station on the Clearwater River (the largest tributary of the Queets River). The variable AnnTemp is a surrogate for annual stream water temperature. Air temperature data were more readily available than stream water temperature data, and the two variables are closely related (Stefan and Preud'homme 1993; Caissie et al. 2001). Water temperature is associated with the survival of spawners and eggs to the smolt stage. Optimal survival of coho salmon eggs occurs at water temperatures in the range 2–8°C (Tang et al. 1987); survival decreases as water temperatures exceed 11°C (Murray and MacPhail 1988).

Winter stream water flow (WinFlow) was calculated as the average of the monthly mean stream flows (ft^3/s; 1 $ft^3 = 28.3$ L) from November prior to the smolt year to March in the smolt year (the second winter from the brood year). Flow data were obtained from the U.S. Geological Survey (http://waterdata.usgs.gov) for the Queets River near its confluence with the Clearwater River.

The sea surface temperature from April to June (SSTAMJ) was calculated as the arithmetic mean of the monthly mean sea surface temperatures from April to June in the smolt year (the primary months of ocean entry for smolts). Sea surface temperature measurements were obtained from the National Climate Data Center (http://iridl.ldeo.columbia.edu/) for the location 48°N, 126°W.

The ocean upwelling index from April to June (UpwellAMJ) was calculated as the arithmetic mean of monthly mean upwelling indices from April to June in the smolt year. Upwelling data were obtained from the Pacific Fisheries Environmental Laboratory (http://www.pfeg.noaa.gov/) for the location 48°N, 125°W.

Sea level pressure from April to June (SLPAMJ) was calculated as the arithmetic mean of the monthly mean sea level pressures (millibars) from April to June in the smolt year. Sea level pressure data were obtained from the National Center for Environmental Protection and the National Center for Atmospheric Research (http://iridl.ldeo.columbia.edu/) for the location 47.5°N, 125°W.

Each environmental variable was standardized by subtracting the mean of all of the values of that variable from each observed value and then dividing that difference by the standard deviation. Standardization facilitated our analyses and minimized the effect of collinearity on the abundance forecasting models in the presence of higher-order polynomial terms (Neter et al. 1996).

Results

The GAM Models Selected

As indicated above, the final forms of models 1B, 2B, 3A, and 3B were selected using nonparametric GAM procedures. For model 1B, a simple linear model

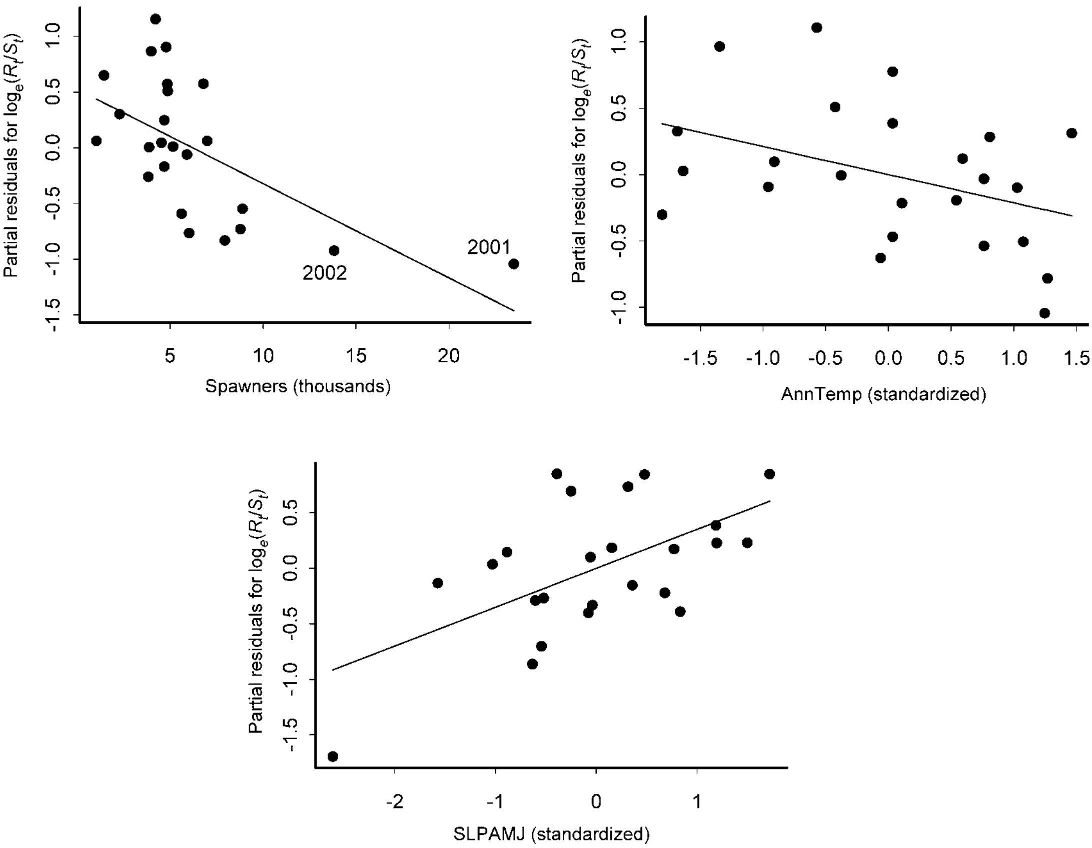

FIGURE 3.—Fit of the GAM model selected for model 2B. The variable AnnTemp is annual air temperature over the 16-month period from December of the brood year to March of the smolt year and serves as a proxy for water temperature. The solid lines represent the component smoothing functions, which are linear for this model. See Figure 2 for additional details.

with SLPAMJ was selected to predict the marine survival rate of Queets River coho salmon (Figure 2). For model 2B, linear functions of AnnTemp and SLPAMJ were identified, along with the number of spawners, to predict adult abundance (Figure 3). For model 3A, a linear function between $\log_e(R_t)$ and the smolt production estimate was identified by the GAM procedure (Figure 4). For model 3B, a linear function of the smolt production estimate and an additional smoothing spline (not linear in this case) of Upwell-AMJ were identified to predict recruits (Figure 5).

For model 2B, the linear relationship of $\log_e(R_t/S_t)$ to S seemed to be driven by the values of S in 2001 and 2002 (Figure 3), but the linear relationship was maintained (not shown) when these two points were removed. We chose to include the data for 2001 and 2002 because they represent brood years with a large number of spawners. In addition, we felt that our ability to decide whether or not an observation is an outlier was limited by the small number of data.

For model 3B, the standardized upwelling index value for brood year 1980 had a strong effect on model selection (Figure 5). Without data for 1980, a smoothing spline of UpwellAMJ with 2 degrees of freedom would have been selected. We chose to include the data for 1980 in the model for reasons similar to those for model 2B.

Parametric Approximations to the Selected GAM Models

A parametric linear or polynomial regression model was formulated for each GAM model based on the functional forms suggested by the nonparametric fits (Table 2). Figure 2 suggests a simple linear regression model between the marine survival rate and SLPAMJ for model 1B. Figure 3 suggests a multiple linear regression model with three predictor variables (the number of spawners, AnnTemp, and SLPAMJ) for model 2B. Figure 4 suggests a simple linear regression model between $\log_e(R_t)$ and the smolt production

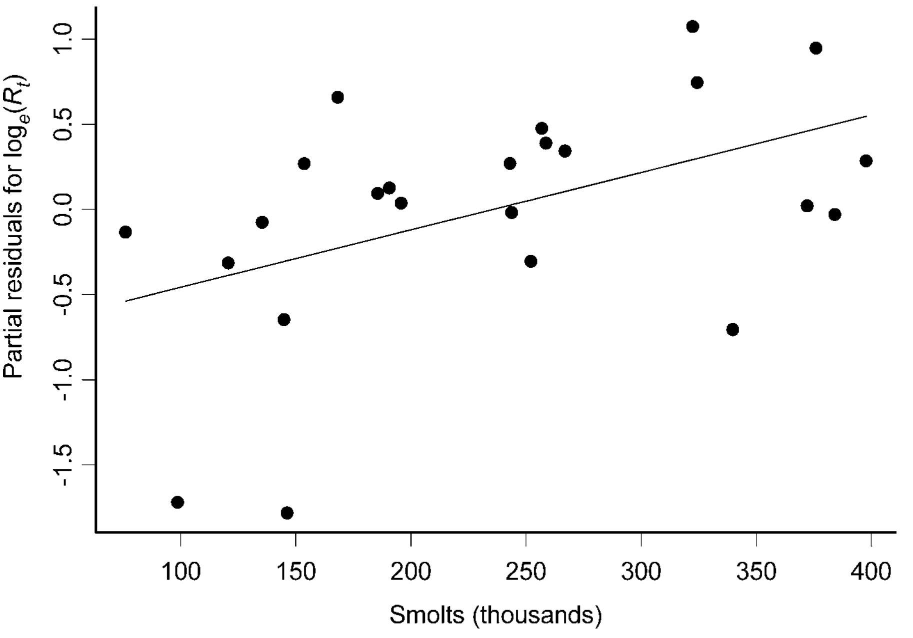

FIGURE 4.—Fit of the GAM model selected for model 3A. The solid line represents the component smoothing function, which is linear for this model. See Figure 2 for additional details.

estimate for model 3A. Figure 5 suggests that a multiple-regression model with the smolt production estimate, UpwellAMJ, UpwellAMJ squared, and UpwellAMJ cubed captures the functional form of the nonparametric model 3B.

Each GAM model selected and its parametric approximation had the same or similar deviance and degrees of freedom (Table 2), suggesting that each parametric model is an adequate approximation to its GAM model (i.e., not significantly different according to the chi-square test).

Performance of the Parametric Models

The parametric forms of the GAM models selected were employed to generate forecasts and evaluate the performance of the alternative forecasting models. The values of the mean absolute percentage error, the largest annual absolute percentage error among all prediction years, and the 90% PI coverage probability for each of the competing parametric models are given in Table 3. The ranks and rank-sums for the performance metrics of competing models are presented in Table 4. When leave-one-out cross-validations were done, model 1B performed the best, and the performance of all models improved with the addition of environmental variables. When hindcasting validations were done, the rank-sum values for models 2B and 3B were the same and lower than those for other models. The addition of environmental variables improved the performance for 2 out of the 3 models.

TABLE 2.—Derivation of the parametric approximations to the GAM models selected for models 1B, 2B, 3A, and 3B. Deviance is defined as the residual sum of squares; variables are listed in Table 1.

	GAM model			Parametric model		
Model	Variables	Approximate residual df	Deviance	Varibles	Residual df	Deviance
1B	SLPAMJ	22	0.01	SLPAMJ	22	0.01
2B	Spawners, AnnTemp, and SLPAMJ	20	5.50	Spawners, AnnTemp, and SLPAMJ	20	5.50
3A	Smolts	22	8.59	Smolts	22	8.59
3B	Smolts and UpwellAMJ[a]	19	3.38	Smolts, UpwellAMJ, UpwellAMJ^2, and UpwellAMJ^3	19	3.37

[a] Smoothing spline with a nominal value of 3 df.

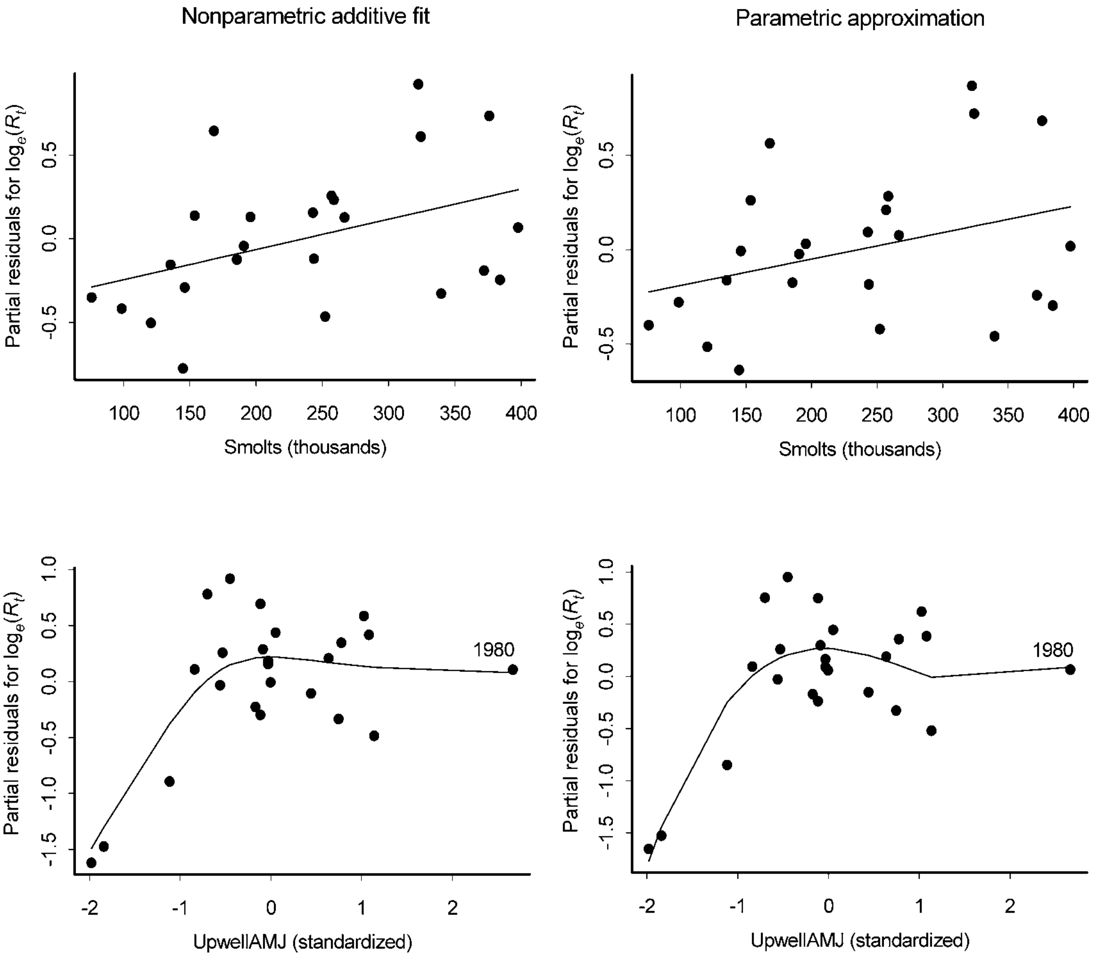

FIGURE 5.—Fits of the nonparametric GAM model selected for model 3B (left column) and its parametric approximation (right column). The nonparametric model has a linear term of the smolt production estimate and a smoothing spline (df = 3) of UpwellAMJ; the parametric model has a linear term of the smolt production estimate and a third-degree polynomial of UpwellAMJ. The variable UpwellAMJ is average monthly upwelling index from April to June during the smolt year. See Figure 2 for additional details.

Discussion

Our study illustrates that the GAM method can be employed to construct forecasting models by identifying promising relationships with the predictor variables. Because functional relationships between the response variable and the predictor variables are determined by the data through analysis, GAM models can be used to identify the best candidates for forecasting models. The GAM methods provide a faster and more objective way of building models than attempting a priori specification of separate models (e.g., linear, exponential, polynomial, etc.) and comparing their goodness of fit. The difficulties in interpreting and applying nonparametric GAM models can be offset by transforming them into parametric forms.

The GAM methods indicate that the predictive performance of the abundance forecasting models for Queets River coho salmon can be improved by incorporating a small number of environmental variables as predictors. The abundance of these fish is correlated with the values of environmental variables that can be readily obtained from monitoring programs. Our analysis shows that annual air temperature (a surrogate for the stream water temperature during incubation and juvenile growth during freshwater residence), sea level pressure (negatively correlated with sea surface temperature, which might have directly affected the survival of coho salmon during

TABLE 3.—Model performance metrics for the leave-one-out cross-validations and hindcasting validations. Abbreviations are as follows: MAPE = the mean absolute percentage error; LAPE = the largest annual absolute percentage error among all years; and 90% PI coverage = the proportion of years for which the observed abundance fell within the 90% prediction interval for the forecast.

Metric	Model 1A	1B	2A	2B	3A	3B
Leave-one-out cross-validations						
MAPE	73.1	47.6	86.8	64.0	77.7	58.2
LAPE	364.5	214.5	784.8	302.3	459.4	414.5
90% PI coverage	0.37	0.29	0.29	0.21	0.42	0.42
Hindcasting validations						
MAPE	81.5	88.8	84.4	77.4	119.3	64.0
LAPE	160.7	234.8	260.0	121.4	431.5	171.3
90% PI coverage	0.20	0.10	0.50	0.30	0.20	0.40

TABLE 4.—Ranks and rank-sums for the performance metrics of competing models. The rank-sum (total) for each model is the sum of the ranks for all performance metrics; lower values indicate better performance. See Table 3 for abbreviations.

Metric	Model 1A	1B	2A	2B	3A	3B
Leave-one-out cross-validations						
MAPE	4	1	6	3	5	2
LAPE	3	1	6	2	5	4
90% PI coverage	3	4.5	4.5	6	1.5	1.5
Total	10	6.5	16.5	11	11.5	7.5
Hindcasting validations						
MAPE	3	5	4	2	6	1
LAPE	2	4	5	1	6	3
90% PI coverage	4.5	6	1	3	4.5	2
Total	9.5	15	10	6	16.5	6

the early months of marine residence; Nickelson 1986; Cole 2000; Koslow et al. 2002; Logerwell et al. 2003), and upwelling (linked with the production of zooplankton, a major source of food for coho salmon during the early months of marine residence; Peterson and Schwing 2003) were associated with the production and survival of Queets River coho salmon.

The average monthly mean stream flow from the November prior to the smolt year to March of the smolt year was not statistically significant in determining recruits per spawner or adult production for Queets River coho salmon in spite of our finding a dome-shaped relationship with smolt production similar to that reported by Lawson et al. (2004). Lawson et al. (2004) interpreted the dome-shaped relationship as indicating that intermediate flows increase smolt production and survival by maximizing presmolt overwintering habitat while high flows reduce smolt production and survival by flushing presmolts out of their refuge areas.

Despite the improvements in abundance forecasting that can result from incorporating environmental variables into predictive models, the absolute percentage errors of the adult abundance forecasts were still substantial for some years (e.g., the large values of LAPE in Table 3). In addition to the measurement errors in spawning escapement, smolt production, or postseason abundance estimates (Ludwig and Walters 1981; Walters and Ludwig 1981), the inability of these models to capture the large variability in the production and survival of Queets River coho salmon under unstable environmental conditions could have caused the large forecasting errors. The limited set of environmental data available may not be sufficient to characterize the multitude of environmental factors that could affect the production and survival of coho salmon. Given spawning habitat that is affected by seasonal flows, an extensive period of rearing in freshwater, a relatively short period of marine entry, and a year and a half of ocean residence, there are many opportunities for complex interactions that could affect production and survival. Additional research should be undertaken to better understand the mechanisms by which environmental factors affect coho salmon production and survival.

Acknowledgments

We thank the personnel of the Quinault Department of Fisheries for collecting the data for our study. We thank John Cornell, Joe Schumacker, Steve Meadows, and Din Chen for assistance with this study. We thank Hap Leon for sharing his study relating fish survival to sea surface temperatures and the Pacific Decadal Oscillation. We also thank Jeremy Fivecrows for help with Figure 1. Finally, we thank Pete Lawson, Saang-Yoon Hyun, and four anonymous referees for reviewing this paper and providing helpful advice.

References

Akaike, H. 1974. A new look at statistical model identification. IEEE Transactions on Automatic Control 19:716–722.

Bradu, D., and Y. Mundlak. 1970. Estimation in lognormal models. Journal of the American Statistical Association 65:198–211.

Caissie, D., N. El-Jabi, and M. G. Satish. 2001. Modelling of maximum daily water temperatures in a small stream using air temperatures. Journal of Hydrology 251:14–28.

Cole, J. 2000. Coastal sea surface temperature and coho salmon production off the northwest United States. Fisheries Oceanography 9:1–16.

Coronado, C., and R. Hilborn. 1998. Spatial and temporal factors affecting survival in coho salmon (*Oncorhynchus kisutch*) in the Pacific Northwest. Canadian Journal of Fisheries and Aquatic Sciences 55:2067–2077.

Duan, N. 1983. Smearing estimate: a nonparameteric retransformation method. Journal of the American Statistical Association 78:605–610.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, New York.

Francis, R. C., and S. R. Hare. 1994. Decadal-scale regime shifts in the large marine ecosystems of the Northeast Pacific: a case for historical science. Fisheries Oceanography 3:279–291.

Hare, S. R., and R. C. Francis. 1995. Climate change and salmon production in the Northeast Pacific Ocean. Canadian Special Publication of Fisheries and Aquatic Sciences 121:357–372.

Hare, S. R., N. J. Mantua, and R. C. Francis. 1999. Inverse production regimes: Alaskan and West Coast salmon. Fisheries 24(1):6–14.

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized additive models. Chapman and Hall, New York.

Hilborn, R., and C. J. Walters. 1992. Quantitative fisheries stock assessment: choice, dynamics, and uncertainty. Chapman and Hall, New York.

Koslow, J. A., A. J. Hobday, and G. W. Boehlert. 2002. Climate variability and marine survival of coho salmon (*Oncorhynchus kisutch*) in the Oregon production area. Fisheries Oceanography 11:65–77.

Lawson, P. W., E. A. Logerwell, N. J. Mantua, R. C. Francis, and V. N. Agostini. 2004. Environmental factors influencing freshwater survival and smolt production in Pacific Northwest coho salmon (*Oncorhynchus kisutch*). Canadian Journal of Fisheries and Aquatic Sciences 613:360–373.

Logerwell, E. A., N. Mantua, P. W. Lawson, R. C. Francis, and V. N. Agostini. 2003. Tracking environmental processes in the coastal zone for understanding and predicting Oregon coho (*Oncorhynchus kisutch*) marine survival. Fisheries Oceanography 12:554–568.

Ludwig, D., and C. J. Walters. 1981. Measurement errors and uncertainty in parameters for stock and recruitment. Canadian Journal of Fisheries and Aquatic Sciences 38:711–720.

Mantua, N., S. Hare, Y. Zhang, J. Wallace, and R. Francis. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the American Meteorological Society 78:1069–1079.

Murray, C. B., and J. D. McPhail. 1988. Effect of incubation temperature on the development of five species of Pacific salmon (*Oncorhynchus*) embryos and alevins. Canadian Journal of Zoology 66:266–273.

Neter, J., M. H. Kutner, C. J. Nachtsheim, and W. Wasserman. 1996. Applied linear statistical models, 4th edition. Irwin, Chicago.

Nickelson, T. E. 1986. Influences of upwelling, ocean temperature, and smolt abundance on marine survival of coho salmon (*Oncorhynchus kisutch*) in the Oregon production area. Canadian Journal of Fisheries and Aquatic Sciences 43:527–535.

Nickelson, T. E., J. D. Rodgers, S. L. Johnson, and M. F. Solazzi. 1992. Seasonal changes in habitat use by juvenile coho salmon (*Oncorhynchus kisutch*) in Oregon coastal streams. Canadian Journal of Fisheries and Aquatic Sciences 49:783–789.

Pearcy, W. G. 1992. Ocean ecology of North Pacific salmonids. University of Washington Press, Seattle.

Peterson, W. T., and F. B. Schwing. 2003. A new climate regime in Northeast Pacific ecosystems. Geophysical Research Letters 30:1896.

PFMC (Pacific Fisheries Management Council). 2006. Review of 2005 ocean salmon fisheries. PFMC, Portland, Oregon.

PSC (Pacific Salmon Commission). 2000. Pacific salmon treaty: 1999 revised annexes, memorandum of understanding (1985), and exchanges of notes. PSC, Vancouver.

Quinn, T. J., and R. B. Deriso. 1999. Quantitative fish dynamics. Oxford University Press, New York.

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Fisheries Research Board of Canada Bulletin 191.

Ryding, K. E., and J. R. Skalski. 1999. Multivariate regression relationships between ocean conditions and early marine survival of coho salmon (*Oncorhynchus kisutch*). Canadian Journal of Fisheries and Aquatic Sciences 56:2374–2384.

Stefan, H. G., and E. B. Preud'homme. 1993. Stream temperature estimation from air temperature. Water Resources Bulletin 29:27–45.

Tang, J., M. D. Bryant, and E. L. Brannon. 1987. Effect of temperature extremes on the mortality and development rates of coho salmon embryos and alevins. Progressive Fish-Culturist 49:167–174.

Venables, W. N., and B. D. Ripley. 1999. Modern applied statistics with S-plus, 3rd edition. Springer-Verlag, New York.

Walters, C. J. 1985. Bias in estimation of functional relationships from time series data. Canadian Journal of Fisheries and Aquatic Sciences 42:147–149.

Walters, C. J. 1990. A partial bias correction factor for stock–recruitment parameter estimation in the presence of autocorrelated environmental effects. Canadian Journal of Fisheries and Aquatic Sciences 47:516–519.

Walters, C. J., and D. Ludwig. 1981. Effects of measurement errors on the assessment of stock–recruitment relationships. Canadian Journal of Fisheries and Aquatic Sciences 38:704–710.