

Evaluating the performance of Gulf of Alaska walleye pollock (*Theragra chalcogramma*) recruitment forecasting models using a Monte Carlo resampling strategy

Yong-Woo Lee, Bernard A. Megrey, and S. Allen Macklin

Abstract: Multiple linear regressions (MLRs), generalized additive models (GAMs), and artificial neural networks (ANNs) were compared as methods to forecast recruitment of Gulf of Alaska walleye pollock (*Theragra chalcogramma*). Each model, based on a conceptual model, was applied to a 41-year time series of recruitment, spawner biomass, and environmental covariates. A subset of the available time series, an in-sample data set consisting of 35 of the 41 data points, was used to fit an environment-dependent recruitment model. Influential covariates were identified through statistical variable selection methods to build the best explanatory recruitment model. An out-of-sample set of six data points was retained for model validation. We tested each model's ability to forecast recruitment by applying them to an out-of-sample data set. For a more robust evaluation of forecast accuracy, models were tested with Monte Carlo resampling trials. The ANNs outperformed the other techniques during the model fitting process. For forecasting, the ANNs were not statistically different from MLRs or GAMs. The results indicated that more complex models tend to be more susceptible to an overparameterization problem. The procedures described in this study show promise for building and testing recruitment forecasting models for other fish species.

Résumé : Nous avons comparé les régressions linéaires multiples (MLRs), les modèles additifs généralisés (GAMs) et les modèles statistiques de réseaux de neurones artificiels (ANNs) comme méthodes pour prédire le recrutement de la goberge de l'Alaska (*Theragra chalcogramma*) dans le golfe de l'Alaska. Nous avons appliqué chaque modèle, basé sur un modèle conceptuel, à une série chronologique de 41 années de données sur le recrutement, la biomasse des reproducteurs et les covariables de l'environnement. Un sous-ensemble de la série chronologique disponible, soit un ensemble de données provenant de l'échantillon consistant en 35 des 41 points de données, a servi à ajuster un modèle de recrutement dépendant de l'environnement. Des méthodes statistiques de sélection des variables ont permis d'identifier les covariables les plus influentes afin de bâtir le meilleur modèle explicatif du recrutement. Un ensemble de six points de données hors de l'échantillon a été gardé pour la validation des modèles. Nous avons testé la capacité de chaque modèle à prédire le recrutement en appliquant le modèle à un ensemble de données hors de l'échantillon. Pour une évaluation plus robuste de la justesse de la prédiction, nous avons testé les modèles avec des essais de rééchantillonnage de Monte Carlo. Les ANNs surpassent les autres techniques durant le processus d'ajustement des modèles. Pour ce qui est de la prédiction, les ANNs ne sont pas statistiquement différents des MLRs et des GAMs. Les résultats indiquent que les modèles plus complexes ont plus tendance à souffrir de problèmes de surparamétrisation. Les procédures que nous décrivons sont prometteuses pour l'élaboration et l'ajustement des modèles de prédiction du recrutement chez d'autres espèces de poissons.

[Traduit par la Rédaction]

Introduction

Society is challenged to steward the exploitation of marine resources effectively for the health of the planet. Forecasting recruitment is a commonly recognized, but elusive, goal for the sustainable stewardship and rational manage-

ment of exploited fish populations. Reliable estimation of recruitment is critically important to assessment of the exploitable segment of those populations (Needle 2001). Furthermore, industry constantly must prepare for changes in abundance, age structure, and spatial distribution of commercial species to optimize economic return and maintain a

Received 7 January 2008. Accepted 27 October 2008. Published on the NRC Research Press Web site at cjfas.nrc.ca on 19 February 2009.

J20353

Y.-W. Lee.^{1,2} Joint Institute for the Study of the Atmosphere and Ocean, Box 354235, University of Washington, Seattle, WA 98195, USA.

B.A. Megrey. NOAA-NMFS, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115-6349, USA.

S.A. Macklin. NOAA-OAR, Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, WA 98115-6349, USA.

¹Corresponding author (e-mail: YongWoo.Lee@dfw.wa.gov).

²Present address: Washington Department of Fish and Wildlife, 600 Capitol Way North, Olympia, WA 98501-1091, USA.

sustainable resource. Managers must make meaningful and effective decisions based on assumptions of the future state of fishery resources. Although the need for accurate recruitment forecasts is great, there are few studies that have focused on this specific problem. An Aquatic Science and Fisheries Abstracts literature search returned only 22 peer-reviewed papers with titles containing the words “recruitment” and either “prediction” or “forecast” (date of search: 4 April 2007). This is a surprisingly small number of papers given the fact that the estimation of recruitment has been a persistent preoccupation of fisheries scientists for more than a century (Cushing 1988; Kendall and Duker 1998).

One source of frustration in predicting recruitment is the seemingly unexplainable complex relationship between fish and their environment (Bailey et al. 2005; Ciannelli et al. 2007). All recruitment models assume that recruitment inevitably will be reduced if there is insufficient spawning biomass (Beverton and Holt 1957; Ricker 1975; Schnute 1985). Often, however, spawner-recruit models fail to adequately reveal processes and mechanisms affecting recruitment and, hence, are consequently unable to predict future recruitment with satisfactory precision. The environmental factors that influence recruitment in a complex, nonlinear, dynamical system often obscure the relationship between spawners and recruits to such a degree that any patterns or relationships, even if present, are not easily identified (Bailey et al. 2005). Furthermore, imprecise measurements of recruitment, parental stock, and environmental factors add additional uncertainty (Walters and Ludwig 1981). We still have only a limited understanding of the recruitment process despite our best efforts to understand recruitment dynamics because of the complex, nonlinear interactions within and between physical and biological variables.

Another obstacle to forecasting recruitment is the lack of established protocols for selecting appropriate statistical approaches. Traditional linear modeling approaches such as multiple linear regressions (MLRs) often have been used to relate external variables to recruitment variability. With the development of powerful personal computers, newer modeling techniques such as artificial neural networks (ANNs) and generalized additive models (GAMs) also have been tried. These new modeling techniques use nonparametric approaches that do not require inflexible assumptions, as opposed to MLRs that require restrictive assumptions such as functional linear relationships between the variables (Chen and Ware 1999; Huse and Ottersen 2003). Megrey et al. (2005) recently explored the performance of MLRs, GAMs, and ANNs to forecast recruitment from simulated data with known properties, but their conclusions were based on one random realization of simulated data.

The goal of this study is to build on the work of Megrey et al. (2005). We intend to advance that exploration by using the same three forecasting techniques and then applying a robust statistical methodology to evaluate forecasting accuracy. The objectives of this paper are (*i*) to build recruitment forecasting models using three different statistical methods, two different response variables, and two model constraints, (*ii*) to fit the models to an in-sample data set of environmental covariates thought to influence recruitment variability of Gulf of Alaska (GOA) walleye pollock (*Theragra chalcog-*

ramma, hereafter referred to as pollock), (*iii*) to use the best-fit models to forecast recruitment by applying them to portions of the time series reserved according to a Monte Carlo resampling strategy, (*iv*) to compare the forecast performance of the models, and (*v*) to recommend forecast modeling techniques based on those comparisons.

Materials and methods

Data sets

We modeled age-2 GOA pollock abundance as a function of spawning biomass and a suite of environmental covariates thought to be influential to recruitment success. At the time of this analysis, estimates of annual recruitment (REC) and spawning stock biomass (SSB) from the annual age-structured stock assessment model (Dorn et al. 2003) were available for the period 1961–2003. We assumed that the recruitment series is not serially correlated or that it is weakly correlated at an ignorable level. This assumption was required because none of the statistical methods used in this study is designed to deal with serial correlations. Also, tests using Monte Carlo resampling are not plausible if there is strong serial correlation. Although no serial correlation was assumed, the degree of serial correlation was examined by employing the sample autocorrelation function on the model residuals at various time lags (1 to 10). The significance of the autocorrelation was assessed with the approximate 95% confidence interval (Brockwell and Davis 2003).

A suite of environmental data was also available for the same period. The recruitment data were lagged 2 years to coincide with SSB and environmental covariates for the birth year, resulting in time series of 41 annual data points spanning the year classes 1961–2001.

Environmental covariates were selected based on a conceptual model of GOA pollock recruitment (Megrey et al. 1996) and the results of an exploratory analysis of the relationship between GOA recruitment success and the physical environment (Megrey et al. 1995). That analysis showed that age-2 recruitment abundance is closely related to precipitation, atmospheric sea-level pressure gradient, and local wind mixing. Guided by this analysis, we chose a subset of these variables as covariates, including local physical parameters, climate-scale indices, and SSB (the only biological variable; variables considered for this study are given in Table 1).

Environmental covariates considered include sea surface temperature (SST), wind mixing energy (WMX), freshwater runoff index (FRN), Northeast Pacific pressure index (NEP), Pacific Decadal Oscillation index (PDO), and Southern Oscillation index (SOI). The environmental data series were obtained as monthly averages. SST and WMX are estimated values, centered on the exit of Shelikof Strait (56°N, 156°W) and derived from the National Center for Environmental Prediction (NCEP) data reanalysis. FRN is an index for integrated GOA coastal freshwater discharge anomaly (Royer 1982). NEP is the sea-level pressure difference between points over the north-central Pacific and near Reno, Nevada (Emery and Hamilton 1985). PDO is the first principal component of the North Pacific monthly SST variability, poleward of 20°N; it describes the decadal variability in

Table 1. Recruitment and explanatory environmental covariates used in this study.

Variable	Acronym	Data source
Age-2 recruitment	REC	Stock assessment estimate (Dorn et al. 2003)
Spawning stock biomass	SSB	Stock assessment estimate (Dorn et al. 2003)
Sea surface temperature	SST	NOAA-NCEP (www.ncep.noaa.gov)
Wind mixing	WMX	NOAA-NCEP (www.ncep.noaa.gov)
Northeast Pacific pressure index	NEP	NOAA-PMEL, extended from Emery and Hamilton (1985)
Freshwater runoff index	FRN	Extended from Royer (1982)
Pacific Decadal Oscillation index	PDO	JISAO (jisao.washington.edu/pdo)
Southern Oscillation index	SOI	NOAA-CPC (www.cpc.noaa.gov)

Note: The descriptions of the variables are given in the text. NOAA, National Oceanic and Atmospheric Administration, US; NCEP, National Center for Environmental Prediction; PMEL, Pacific Marine Environmental Laboratory; JISAO, Joint Institute for the Study of the Atmosphere and Ocean, University of Washington; CPC, Climate Prediction Center.

cool and warm phases of Pacific environmental regimes (Mantua and Hare 2002). SOI, the anomaly in the sea-level pressure difference between Tahiti (18°S, 150°W) and Darwin (10°S, 130°E), is a good indicator of tropical variations related to El Niño events (Trenberth 1984). Although biological variables such as predation and prey availability are known to affect recruitment success, the only biological variable considered for the forecasting models in this study was SSB. Other available biological time series were incomplete, short in length, and did not overlap the recruitment time series enough to make them useful.

Environmental covariates considered in this study have different temporal and spatial scales of influence. Some are regional and others are basin scale in their spatial scope. Similarly, in the temporal domain, some factors are important in establishing optimum conditions prior to spawning, some during spawning and larval life stages, and some during early juvenile stages. In building recruitment forecast models, we averaged the monthly environmental covariate data over 3-month periods considered to be important to pollock recruitment. These correspond to pollock prespawning (January–March), spawning (April–June), and early juvenile (July–September) life history periods. We elected to ignore the period from October through December, by which time young-of-the-year pollock are independent free-swimmers and, assuming they have reached critical size to survive the upcoming winter, are much less susceptible to the environmental covariates we chose (Bailey 1989). We treated the average of each 3-month period for each environmental covariate as a separate explanatory variable for the model. Thus, there were 19 covariates, including SSB, available for analysis.

For the purpose of identification and to clarify the presentation, we added to the end of each environmental covariate's acronym a number (1, 2, or 3) that describes the temporal influence of the environmental covariate on the life history period (prespawning, spawning, or early juvenile, respectively). All environmental time series covariates were normalized $((x - \mu_x) / \sigma_x)$ prior to analysis (Fig. 1).

The recruitment forecast of GOA pollock is currently made at five categorically ordered levels of recruitment strengths (weak to strong) based on a weighted scoring of the assemblage of different biological and physical information in the region (Dorn et al. 2003). The recruitment forecast is used to project the future stock status, consequently to recommend the fishing quota to the fisheries managers.

This is the reason that recruitment forecasts remain a vital information component required by resource management decision-makers dealing with exploited marine ecosystems. In this study, the recruitment forecast is modeled and tested using an abundance scale, rather than subjective ordinal scale, thus providing more information to fisheries managers.

Environment-dependent spawner-recruit model

We adopted the generalized Ricker (1975) spawner-recruit model to specify recruitment as a function of spawning biomass and other environmental covariates as generalized by Hilborn and Walters (1992) to include environmental covariates.

$$(1) \quad REC = \alpha SSB \exp(-\beta SSB + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_m X_m) \exp(\omega)$$

where REC is age-2 recruitment (billions of recruits), SSB is the spawning stock biomass (1000 tonnes (t)), X_m is an influential environmental covariate, m is the number of environmental covariates, α , β , and γ_m are model parameters to be estimated, and ω is a normally distributed random error ($\sim N(0, \sigma^2)$).

A natural logarithmic transformation of eq. 1 normalizes the multiplicative lognormal error structure to an additive normal error and linearizes the relationship between recruitment and environmental covariates. From this point on, "log" refers to the natural logarithm "ln".

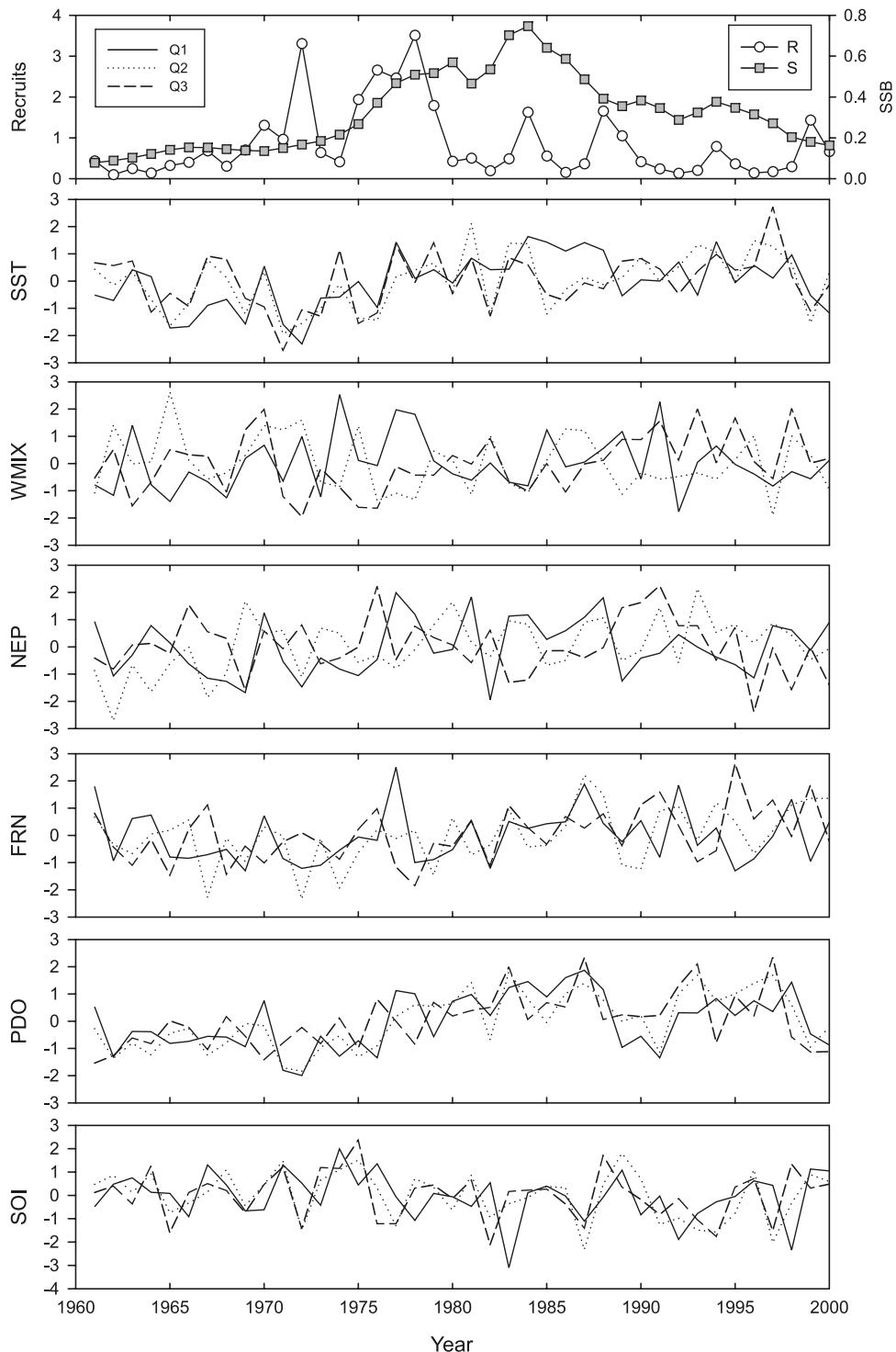
$$(2) \quad \log(REC) = \log(\alpha) + \log(SSB) - \beta SSB + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_m X_m + \omega$$

Equation 2 is further simplified by moving $\log(SSB)$ to the left side of the equation. This models $\log(\text{survival})$ as a function of a mean response and the influence of the environmental covariates.

$$(3) \quad \log(REC/SSB) = \log(\alpha) - \beta SSB + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_m X_m + \omega$$

It should be noted that $\log(SSB)$ in eq. 2 does not have an associated parameter. Thus, representing $\log(\text{survival})$, i.e., $\log(\text{REC}/\text{SSB})$, as the response variable as in eq. 3 is a theoretically preferable form of modeling recruitment in a linear regression setting (Ricker 1975). However, if we do not restrict ourselves to the theoretical Ricker model, we can set

Fig. 1. Gulf of Alaska (GOA) walleye pollock (*Theragra chalcogramma*) recruitment (open circle, billions) and spawning biomass (shaded square, million tonnes) time series, along with time series of normalized environmental covariates used for building GOA pollock recruitment forecasting models in this study. The environmental covariates are 3-month averages of monthly data corresponding to pollock spawning life history: Q1, prespawning, January–March; Q2, spawning, April–June; Q3, early juvenile, July–September.



up a log(recruitment) model as a function of SSB and other environmental covariates as follows:

$$(4) \quad \log(\text{REC}) = \log(\alpha)' - \beta \text{SSB} + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_m X_m + \omega$$

For developing our recruitment forecast models, we con-

sidered two different types of response (dependent) variables: log(recruitment) (eq. 4) and log(survival) (eq. 3).

Statistical methods

We examined three statistical modeling methods (MLRs, GAMs, ANNs) to evaluate their performance at data fitting

and forecasting recruitment. We chose these modeling techniques because GAMs and ANNs have recently received more attention from fisheries scientists as alternatives to traditional linear regression methods for recruitment data analysis and recruitment forecasts (Chen and Ware 1999; Daskalov 1999; Huse and Ottersen 2003). These statistical modeling techniques were used to fit the GOA age-2 pollock recruitment data with SSB and the environmental covariates using models of eqs. 3 and 4.

MLR is a standard way of fitting a linearized form of environment-dependent spawner–recruit models (Hilborn and Walters 1992). However, MLR requires restrictive a priori assumptions of linearity in the functional relationships between the response and the explanatory variables, normality of the underlying error distribution, and homogeneity of variances (Draper and Smith 1998). Because of the highly variable nature of fish recruitment data, it is often difficult to justify linear regression as the appropriate method for the analyses. Despite these shortcomings, linear regression models are still widely used in the analyses of fish recruitment (Wespestad et al. 2000; Rosenkranz et al. 2001; Meuter et al. 2006). The merits of linear regression arise from the fact that the method is conceptually straightforward and has a solid theoretical background for parameter estimation and hypotheses tests.

The GAMs model the data nonparametrically by using scatterplot data smoothers, allowing the GAMs to suggest the functional relationship between response and explanatory variables (Hastie and Tibshirani 1990; Chambers and Hastie 1992) without the need for restrictive a priori assumptions on the functional relationship between the two. There are a number of examples of GAMs used in fisheries applications (e.g., Swartzman et al. 1995; Beare and Reid 2002; Porter et al. 2005). The usefulness of GAMs in the modeling of environment-dependent spawner–recruit models is demonstrated by the work of Jacobson and MacCall (1995), Daskalov (1999), and Cardinale and Arrhenius (2000). GAMs assume that each explanatory variable affects the response variable in a smooth, continuous way. The amount of smoothing is determined by the number of degrees of freedom applied to the smoothing spline function applied to each explanatory variable. We used cubic spline scatterplot smoothers in our application of the GAMs to the recruitment forecast problem.

ANNs are machine-learning algorithms designed to mimic the operation of neurons in the human brain and its processing of input signals for learning and recognizing complex patterns. The human brain is a highly complex and nonlinear information-processing system with interconnected neurons, functioning like a set of numerous parallel computers working simultaneously. Like GAMs, ANNs are useful in situations where there is little information about the mathematical form of the relationship between response and explanatory variables or when the system under study is highly nonlinear, conditions that almost perfectly describe the typical situation of trying to fit fisheries recruitment models. However, ANNs are often criticized or referred to as black boxes because the parameters in the fitted networks lack biological interpretation. The use of ANNs in aquatic and fisheries science has been growing (e.g., Akoi and Komatatsu 1997; Huse and Gjøsæter 1999; Laë et al. 1999), and

the approach has demonstrated promising utility for recruitment forecasts (e.g., Chen and Ware 1999; Huse and Ottersen 2003). There are many different neural network architectures with different training algorithms. For this study, we employed a simple but widely used ANN, a multi-layer feed-forward neural network with a back-propagation learning algorithm (Ripley 1996; Venables and Ripley 2002).

The fundamental difference between ANNs and conventional parametric statistical methods such as nonlinear regressions or multiple linear regressions is in the existence of hidden neurons in the hidden layer for ANNs. ANNs with no hidden neurons are essentially identical to nonlinear regression or linear regression (Sarle 1994). The number of hidden neurons in ANNs affects the accuracy of the models' fit to the data. In general, an ANN model with more hidden neurons (i.e., more parameters) will fit the data better, similar to using higher-order polynomial equations to describe curvilinear response functions. ANN models are so flexible that they can approximate virtually any function to any degree of accuracy given sufficient hidden neurons. However, similar to common curve-fitting techniques, if there are too many hidden neurons in the network, the model tends to fit to the noise and not the underlying patterns in the data. Consequently, forecasting performance for unseen data decreases because of overparameterization, and the statistical model's degrees of freedom quickly disappear.

Fitting and forecasting performance of ANNs are often a zero-sum game requiring a careful balancing act. With the overparameterization problem always looming, there are no established protocols for determining the optimum number of hidden neurons to balance fitting performance with forecasting performance. Yet, because of the heavy influence of hidden neurons on the fitting and forecasting performance of an ANN model, it still remains clear that it is critically important to carefully decide the number of hidden neurons to be used in ANN models. Because of these problems, selecting the number of hidden neurons is usually a difficult subjective decision made by the analyst. We offer a possible objective solution to this dilemma by proposing a systematic increase in hidden neurons and a simultaneous evaluation of the fitting and forecasting performance of the model.

Model development

Model development entailed selection of the X_m environmental covariates and estimation of the model parameters to best fit the recruitment time series. For initial development, we used the first 35 points of the recruitment and covariate time series to establish their functional relationship. Later, in a more robust simulation, we used the same environmental covariates selected in initial model development, but provided 300 separate recruitment realizations, as described later in the Monte Carlo resampling section, to tune the covariate coefficients.

Various statistical procedures are available for selecting explanatory variables for each of the three types of models. Generally, including more covariates in a model improves the fit. However, good performance in model fitting does not always result in good performance in forecasting unseen data. Often, the models forecasting performance is diminished by including too many covariates. This is called over-

parameterization or overfit in statistics (Burnham and Anderson 2002), and it should be avoided in model development. Redundant and unnecessary explanatory variables should also be avoided because they produce less precise inferences about estimated coefficients of the covariates. For these reasons, we adopted the “principle of parsimony” (the scientific principle that things are usually connected or behave in the simplest or most economical way) in building the final version of each statistical model (Burnham and Anderson 2002).

To determine the best MLR model, we used four procedures to select covariates: stepwise selection, Mallow’s C_p statistic, Akaike’s information criteria (AIC), and Swartz’s Bayesian information criteria (BIC). Stepwise procedures allow variables to be added or deleted from the model at each step at a given level of statistical significance (Rawlings et al. 2005). Mallow’s C_p , AIC, and BIC include a penalty term in the formulation of their objective function that is a positive function of the number of parameters in the model. These metrics use this technique to avoid or minimize over-parameterization (Mallows 1973; Schwarz 1978; Akaike 1981). Each procedure returns a ranked order of candidate models. We examined the top five model candidates from each selection procedure and chose the final MLR model that ranked the highest across the four procedures.

Selection of the final GAM model was based on the AIC criterion (Chambers and Hastie 1992). Suitable GAM models were identified by a stepwise selection procedure in which the smoothing function for each explanatory variable was started with one degree of freedom and incremented to a given maximum number of degrees of freedom. We used a maximum of four degrees of freedom for each covariate during the selection procedure.

There is no formal way of selecting explanatory variables for ANN models. The ANN complex model structure contains input, hidden, and output neurons interconnected in a network structure, thus the number of parameters increases nonlinearly as a function of the number of hidden neurons, and these models quickly exhaust degrees of freedom as more covariates are introduced. This is an important consideration in data-limited situations, such as this study with its short recruitment time series. We used the same covariates selected by the final MLR model in the ANN model, because MLRs are equivalent to ANNs with no hidden neurons. Given the set of explanatory variables, we determined the optimum number of hidden neurons by fitting the data set with a variable range of hidden neurons, ranging from one to four. Each ANN model can converge to a different result due to the effect of random initialization of weights. To overcome this difficulty when we were searching for the optimum number of hidden neurons, we used 100 different runs with different random seeds and averaged the results as our output for the final ANN model.

In the variable selection process, it is difficult to establish if recruitment forecasting models should contain SSB as a covariate in the models of eqs. 3 and 4. Often SSB is included even when it is not statistically significant, because some argue that including SSB in the model is necessary for biological realism — there can be no recruits without spawners. To address this uncertainty, we examined the outcome of each final model under two conditions. In the first,

we accepted covariates chosen by the statistical selection procedures. Under the second condition, we “constrained” SSB to be included as an explanatory variable throughout the selection process. This second set of models presumes significant density dependence on recruitment. Constraining SSB to be included in the model during the variable selection procedure can result in a different suite of covariates for the final model than in the nonconstrained condition.

In all, 12 different model configurations were developed based on statistical method (MLR, GAM, ANN), response variable type (log(recruitment), log(survival)), and constraint on SSB during the selection process (no, yes). For an efficient description of different model types, a model identification (model ID) nomenclature was developed using the following rules. In the model ID, the first letter indicates statistical method (M for MLR, G for GAM, A for ANN), and the second letter indicates response variable type (R for log(recruitment), S for log(survival)). If SSB was constrained to be included in the model, a letter C is placed in the third position. If the model was not forced to use SSB, then the third letter is left blank. For example, model GRC is a GAM predicting log(recruitment) with SSB forced to be a covariate (Table 2 lists the various models and their properties).

Forecast evaluation

Most predictive modelling studies use part of the data set to create models and another part to explicitly test the predictive power of the models. In our study, forecast accuracy was measured by calculating a goodness-of-fit metric in a two-step process. First, a subset of available observations, usually called the in-sample set, is used to fit a statistical model, then the fitted model is given a hidden set of data (the out-of-sample set) that is excluded from model fitting. In ANN applications, these are called the training set or segment (in-sample data set) and the forecasting set or segment (out-of-sample data set). Although previous studies showed the promising possibility of using nonparametric modeling techniques to forecast, comparisons of forecasting performance were limited to the last few observations in the data time series or a few subsets of the data series.

The mean absolute error (MAE) was used to measure how well each model fit the in-sample (training) data and how well it forecasted the out-of-sample (forecasting) data:

$$\text{MAE} = \sum_{i=1}^n |Y_i - \hat{Y}|/n$$

MAE is an average of absolute difference between observed and predicted (or forecasted) values. There is an inverse relationship between MAE and prediction accuracy, i.e., MAE decreases as prediction accuracy improves.

Initial method

Recruitment and covariate time series were split into an in-sample (training) set and an out-of-sample (forecasting) set. It is standard practice to reserve the end portion of a time series for forecast evaluation. The exact number of data points to reserve for forecasting is a somewhat arbitrary decision. Initially, an in-sample data set of the first 35 data points from the time series was used to fit the functional re-

Table 2. Final models from the variable selection procedures.

Model								
ID	Method	Response	C*	df	r_1	Covariates	Coefficients	$p (> t)$
MR	MLR	$\log(R)$	No	3	<u>0.398</u>	WMX1 WMX3	0.3381 -0.3033	0.0209 0.0438
MRC	MLR	$\log(R)$	Yes	6	0.219	SSB SST1 WMX1 NEP1 PDO3	3.4496 -0.5712 0.3107 0.2891 -0.4668	0.0060 0.0128 0.0263 0.0970 0.0154
MS	MLR	$\log(R/S)$	No	6	0.182	SST1 WMX1 WMX3 NEP1 PDO3	-0.5310 0.2712 -0.235 0.3023 -0.478	0.0025 0.0278 0.0602 0.0527 0.0014
MSC	MLR	$\log(R/S)$	Yes	7	0.175	SSB SST1 WMX1 WMX3 NEP1 PDO3	0.0819 -0.5394 0.2702 -0.2333 0.3026 -0.4853	0.9392 0.0102 0.0322 0.0708 0.0568 0.0072
GR	GAM	$\log(R)$	No	3	<u>0.418</u>	s(WMX1, 2) [†] , WMX3		
GRC	GAM	$\log(R)$	Yes	7	<u>0.233</u>	SSB, SST1, s(WMX1, 2) [†] , NEP1, PDO3		
GS	GAM	$\log(R/S)$	No	4	0.327	s(SST1, 2) [†] , PDO3		
GSC	GAM	$\log(R/S)$	Yes	5	<u>0.366</u>	SSB, s(WMX3, 2) [†] , PDO3		
AR	ANN	$\log(R)$	No	3	<u>0.384</u>	WMX1, WMX3		
ARC	ANN	$\log(R)$	Yes	7	<u>0.136</u>	SSB, SST1, WMX1, NEP1, PDO3		
AS	ANN	$\log(R/S)$	No	4	0.140	SST1, WMX1, WMX3, NEP1, PDO3		
ASC	ANN	$\log(R/S)$	Yes	5	0.056	SSB, SST1, WMX1, WMX3, NEP1, PDO3		

Note: Entries include model, statistical method, response variable (R, recruitment; S, survival), constraint (presence or absence), df (degrees of freedom), serial autocorrelation at time lag 1 (r_1), selected explanatory covariates, estimated coefficients, and p values. Significant autocorrelations, tested based on the approximate 95% confidence interval, are underlined.

*C denotes the constraint on spawning stock biomass (SSB). If the constraint is given (yes), the SSB is forced to be included in the model and remains during the variable selection process.

[†]Number in parentheses indicates the degrees of freedom for GAM spline smoothers.

lationship of the covariates with recruitment to develop forecasting models, and the remaining six out-of-sample data points were reserved to assess the forecasting skill of the models. This evaluation technique provided a single forecast accuracy assessment for each of the three models in their various forms.

Monte Carlo resampling

Having a single out-of-sample data set for each model does not allow statistical evaluation of the forecasting performance. We attempted a more robust evaluation by using a Monte Carlo statistical resampling strategy to identify data sets subjectively for forecasting. A number of observations were randomly selected and reserved before fitting the forecast models. Later, those reserved observations were compared with the forecasts from the fitted recruitment models. Repeating the process of Monte Carlo random sampling and refitting the forecast models resulted in a set of indices suitable for measuring forecast accuracy.

We reserved six data points for the out-of-sample data set with the intention of including two observations for each of

three qualitative recruitment strength categories. The three categories were determined based on equal percentile partitions of the entire recruitment time series of 41 observations: high (100%–66.7%), middle (66.6%–33.4%), and low (33.3%–0%). Two data points from each recruitment category were selected randomly, yielding the out-of-sample data set consisting of six data points for forecast testing and model validation. The corresponding environmental covariates were reserved as well. The remaining 35 in-sample data points were retained for re-estimating the parameters of the recruitment model using the environmental covariates identified earlier. Three hundred pairs (in-sample and out-of-sample) of resampled data sets were generated using the above procedure. Generation of 300 data sets is somewhat arbitrary, but we felt that 300 replicates were sufficient to compare the forecasting accuracy among the models statistically. A forecast was made based on each refitted model for the corresponding portion of out-of-sample (reserved) data. This procedure was repeated for all 300 pairs of resampled data sets.

Sampling the out-of-sample data set from different re-

ruitment categories ensures that equal numbers of different levels of recruitment strength are contained in a given out-of-sample data set. This sampling scheme was chosen in an attempt to cover the spectrum of recruitment variability as much as possible for a given out-of-sample data set. Although not employed in this study, alternatively, it is possible to sample the out-of-sample data with different sampling schemes, e.g., total random selection regardless of recruitment level or selection from only a particular recruitment level.

A multifactor analysis of variance (ANOVA) was applied to the 300 Monte Carlo coefficients of determination of the fitting and forecasting results. Three factors were examined statistically to test for significant influence on fitting and forecasting performance: statistical method (MLR, GAM, ANN), response variable ($\log(\text{recruitment})$, $\log(\text{survival})$), and constraint on SSB during the selection process (no, yes). Because ANOVA only reveals if there is a difference in the means and not which means are different from each other, a Student–Neuman–Keuls (SNK) multiple comparison test was used to compare the means of MAE values of model performance at a significance level of 0.05.

Results

Each statistical model selected different explanatory variables to fit to the response variables ($\log(\text{recruitment})$ and $\log(\text{survival})$) (Table 2). Selected covariates for the $\log(\text{recruitment})$ models were identical between the MLR and GAM models, except the GAM allowed flexible, varying degrees of freedom for WMX1. More explanatory variables were selected when SSB was constrained to be in a model. When $\log(\text{survival})$ was modeled, MLR selected more covariates compared with the $\log(\text{recruitment})$ model. Compared with the unconstrained MLR $\log(\text{survival})$ model, constraining SSB did not alter the set of explanatory variables. However, the SSB covariate in the constrained model was highly insignificant ($p = 0.9392$), indicating that adding this term to the model did not significantly improve the fit. The GAM technique resulted in different sets of explanatory variables for the constrained and unconstrained models for both response variables ($\log(\text{recruitment})$ and $\log(\text{survival})$). Constrained SSB models had more covariates for GAM than for MLR. Out of 19 possible explanatory variables, the maximum number of covariates selected for the final models was six (MLR for $\log(\text{survival})$ with constraint). The examination of serial correlations in the residuals of all 12 models showed that there were four significant serial correlations at time lag 1 (Table 2). However, we judged that the magnitudes of the correlations (maximum = 0.481) were not substantial enough to seriously violate the assumption of no serial correlation. None of the serial correlations was significant beyond time lag 1.

Interestingly, SSB was never selected as an important covariate in any final model. However, in every case, constraining SSB to be in the model during the variable selection procedure resulted in richer models with more explanatory variables. GAM selected fewer explanatory variables than MLR for $\log(\text{survival})$. Another interesting finding in the variable selection procedure was that no spawning season (season number 2) covariates were chosen by any of the final models.

Several consistent patterns emerged. In all but one model (GS), wind mixing was selected as an important covariate. Wind mixing prior to spawning (WMX1) in the MLR always had a positive contribution to recruitment, and the magnitude of its coefficient was fairly uniform (~0.3). WMX3 had a similar, but opposite, impact. Whenever WMX1 appeared in a GAM model (GR and GRC), it was either concave or asymptotic (Fig. 2). WMX3, on the other hand, was negative and linear in GR and was convex in GS.

The optimal number of ANN hidden neurons was determined by systematically fitting the final ANN models, varying the number of hidden neurons, and comparing the models' performance. For the in-sample data set, the fit improved steadily for AR and AS models, with MAE approaching zero as more hidden neurons were added (Fig. 3). For the out-of-sample data set, however, the forecast did not improve by adding more hidden neurons to the model. Forecasting accuracy became worse as more neurons were added, indicating possible overparameterization. From the result of this experiment, we decided to use one hidden neuron for the final ANN models because more than one hidden neuron in the model appeared to cause overparameterization. We calculated the MAE values of the final models and compared in-sample and out-of-sample performance (Fig. 4). For the in-sample data set, constrained ANN models performed better (smaller MAE) than all other models. GAMs performed better than MLRs for $\log(\text{recruitment})$ but not for $\log(\text{survival})$. In general, $\log(\text{survival})$ models fit the in-sample data set better than $\log(\text{recruitment})$ models. This result can be attributed to the greater number of explanatory variables in the $\log(\text{survival})$ models.

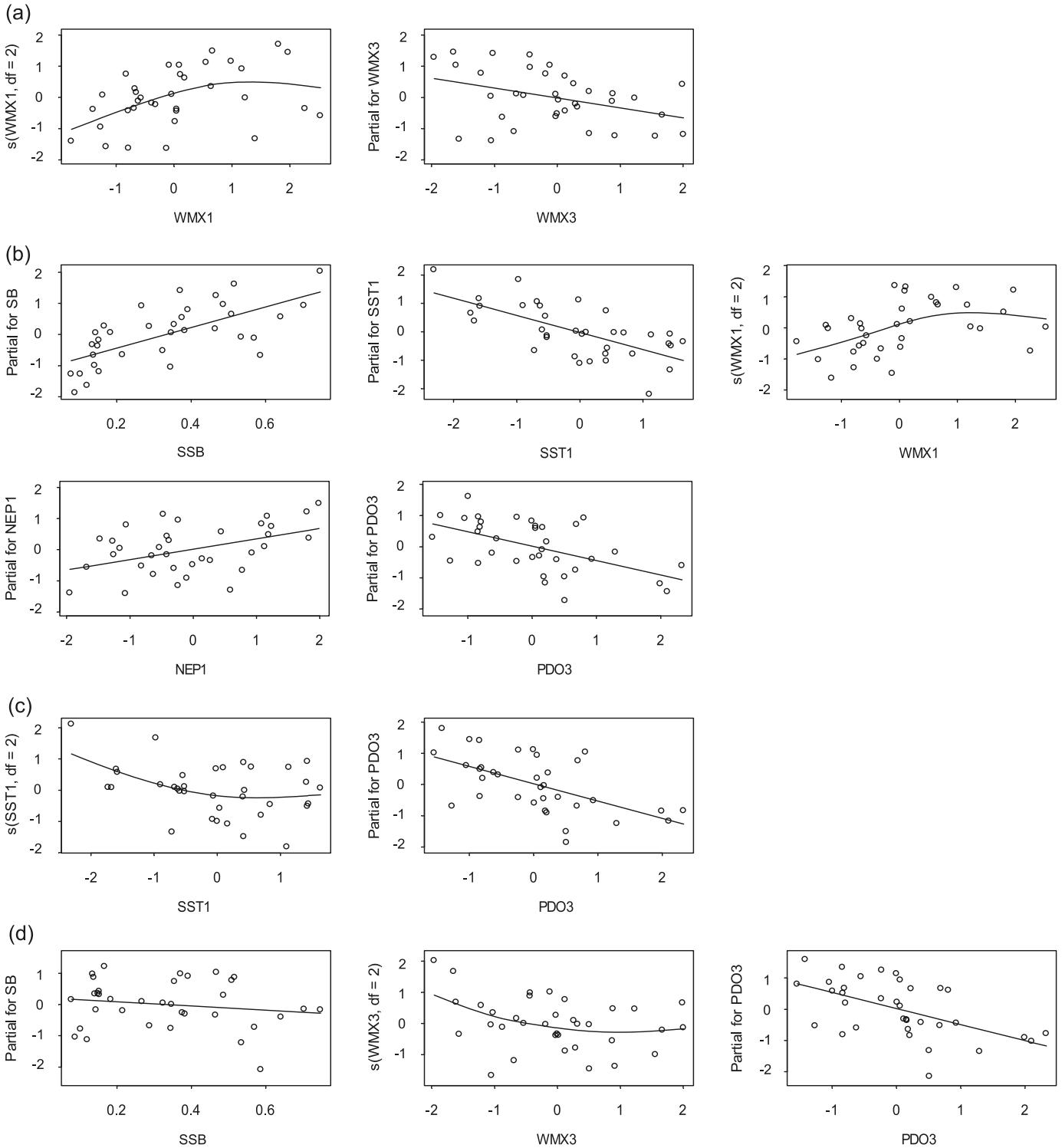
The fitted models were applied to the out-of-sample data set reserved from the end of the time series to test their forecasting performance. Using MAE as the performance measure, MLR forecasting performance was highest across all model cases, except MR (Fig. 4). When constrained and applied to $\log(\text{recruitment})$, MRC (0.49) and GRC (0.55) had higher accuracy compared with $\log(\text{survival})$. ASC (0.59) outperformed ARC (0.65). All unconstrained $\log(\text{recruitment})$ models forecast poorly ($\text{MAE} > 0.79$).

The forecast performances just described cannot be compared statistically because fitting and forecasting were applied to only one data set. To overcome this shortcoming, the performance of all models was compared statistically by creating 300 data sets using the previously described Monte Carlo resampling strategy. Mean MAE values, with 95% confidence, compared fitting and forecasting performances of the models applied to the resampled data sets (Fig. 5).

In general, model fitting was so accurate that confidence intervals cannot even be seen in Fig. 5a. An ANOVA (not presented) of fitting performances based on the 300 resampled data sets showed that all three factors (response, method, and constraint) and their interactions were statistically significant ($p < 0.05$).

The MAE values for forecasting performance had much wider confidence intervals (Fig. 5b). At first glance, no single statistical method stood out as superior, although $\log(\text{survival})$ models appeared to perform better in general. GAMs performed best for $\log(\text{recruitment})$, and MLRs performed best for $\log(\text{survival})$. The forecasting performances of ANNs were disappointing — lowest for $\log(\text{recruitment})$

Fig. 2. Results of the application of GAM models in the following configurations as described in Table 2: (a) GR, (b) GRC, (c) GS, and (d) GSC. Partial residuals for the fitted line of the variables are plotted. Smoothing amount and degrees of freedom for the smoothers of the fitted line of the variable are indicated in the Y axis label, if other than linear.



and moderate for log(survival). An ANOVA (Table 3) revealed that the main factors (response, method, and constraint) and some of their interactions were statistically significant for forecasting performance. The choice of response variable was the most influential factor for forecasting performance based on the relative importance gauged by

the mean square error (MS). Among the main factors, including the SSB constraint was more influential for forecasting performance than the choice of statistical method. Significant interactions between main factors indicated that one factor did not dominate the determination of forecasting performance. Interactions between response variable and

Fig. 3. Effect on model performance of the number of ANN model hidden neurons as measured by MAE values. Solid symbols represent the in-sample fitting set, open symbols, the out-of-sample forecasting set. Solid lines designate results from models with log(recruitment), and broken lines indicate results for log(survival) models. Circles are for models with no constraint on SSB during variable selection, and triangles are for models with constraint on SSB.

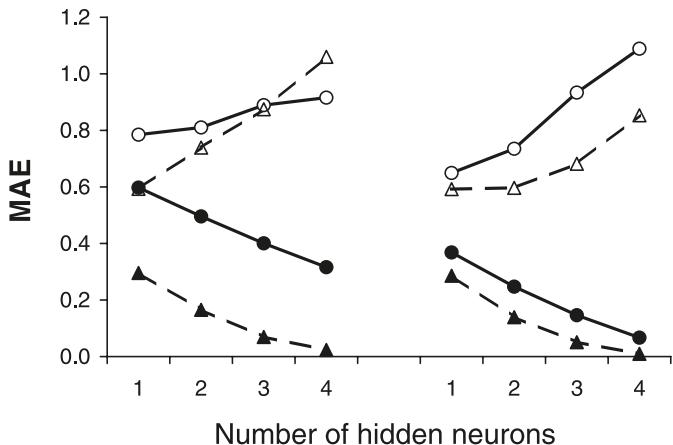
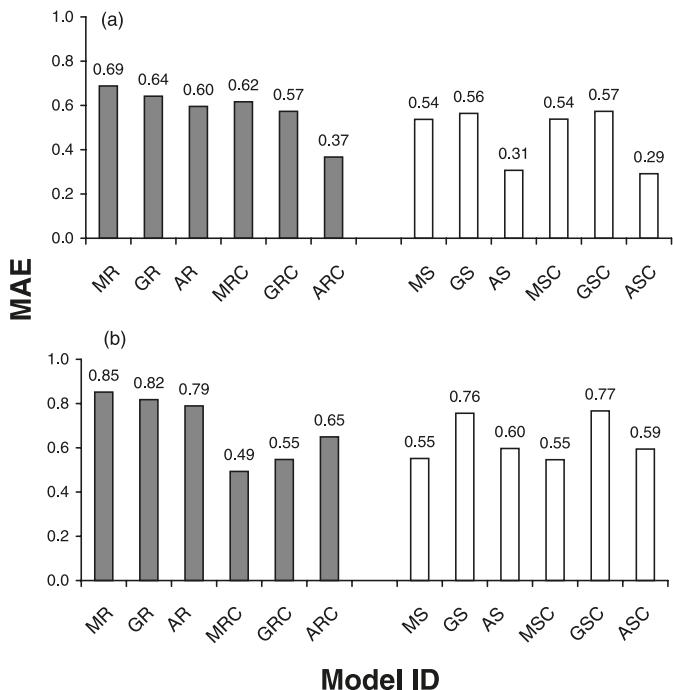
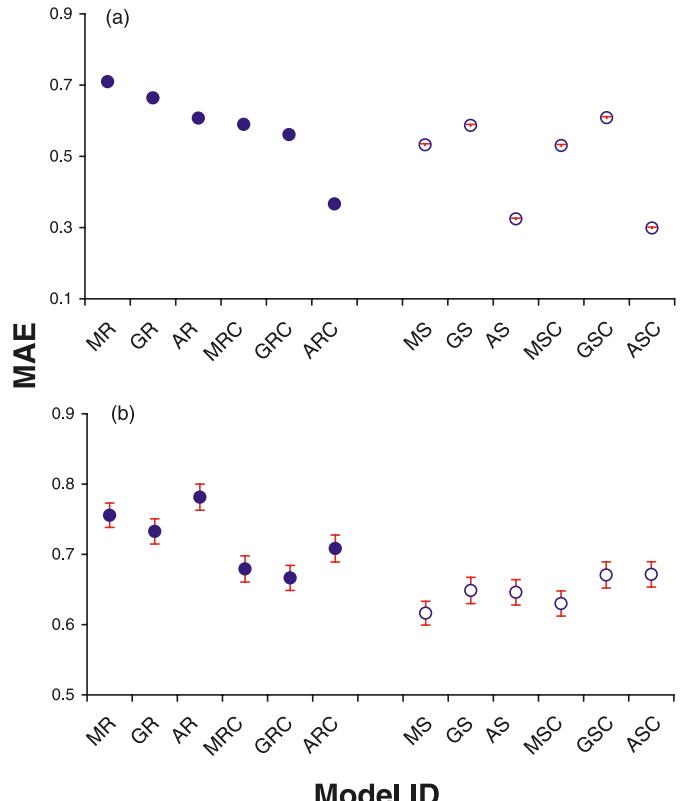


Fig. 4. Comparison of MAE values of various final MLR, GAM, and ANN models described in Table 2 for the in-sample fitting set (a) and the out-of-sample forecasting set (b). Shaded bars represent MAE for models with log(recruitment); open bars show MAE for models with log(survival).



statistical method were significant because the performance of the statistical methods changed with the choice of response variable; GAMs and MLRs were the best statistical method when log(recruitment) and log(survival) were the response variables, respectively. Significant interactions between response variables and constraint factors demonstrated that forcing the SSB covariate to be included

Fig. 5. Performance measures (MAE) of final response variables models when applied to 300 resampled data sets and compared between different statistical models as described in Table 2 for the in-sample fitting sets (a) and the out-of-sample forecasting sets (b). Shaded circles indicate MAE for models with log(recruitment); open circles indicate MAE for models with log(survival). Error bars on the circles indicate ± 2 SE (standard error). For the fitting sets, error bars are not easily visible because of the models' high accuracy at fitting.



in the model during the variable selection procedure had a positive, significant effect on forecasting in one response variable case (log(recruitment)) but not in the other case (log(survival)).

Given that modeling with log(survival) as the response variable resulted in better forecasts, we compared the forecasting performance of the log(survival) models by applying the SNK multiple comparison test (Table 4). There was no statistical difference for the forecasting performance of all three SSB-unconstrained models (MLRs, GAMs, and ANNs), although MLRs had lower observed MAE averages. The forecast performances of ANNs and GAMs became poor when models were constrained with SSB and their MAE means were statistically different from MLRs means. Results also indicated that imposing the SSB constraint consistently reduced the forecasting performance within a statistical method, but the differences were not statistically significant.

Discussion

The recruitment models constructed in this study selected a suite of environmental covariates describing underlying recruitment variability, despite the lack of biological covari-

Table 3. Results of analysis of variance (ANOVA) on the MAE values of all forecasting models for the resampled out-of-sample data sets.

Source	SS	df	MS	F	p (> F)
Method	0.628	2	0.314	12.2	<0.0001
Response	4.854	1	4.854	187.8	<0.0001
Constraint	0.595	1	0.595	23.0	<0.0001
Method × response	0.515	2	0.257	10.0	<0.0001
Method × constraint	0.015	2	0.007	0.3	0.7531
Response × constraint	1.919	1	1.919	74.2	<0.0001
Method × response × constraint	0.005	2	0.002	0.1	0.9160

Note: SS, sum of squares; df, degrees of freedom; MS, mean square error; F, F statistics.

ates. Among the three statistical methods, ANNs performed noticeably well for the in-sample data set, achieving close to a perfect fit to observed data as more hidden neurons were used. The high performance of ANNs in model fitting indicates the high flexibility of ANNs as function approximators. However, the high performance achieved in fitting was not duplicated when the fitted model was applied to the unseen out-of-sample data set for forecasting. This tension between fitting and forecasting performance as more hidden neurons were introduced suggests that ANNs can easily be susceptible to overparameterization when used to model fisheries recruitment time series of limited length. This is exactly the situation that we face with the GOA pollock recruitment time series, even though the data set is relatively long compared with the recruitment time series of other fish species in the world (Myers et al. 1995). If the GOA pollock recruitment time series is sensitive to overparameterization by ANNs, then shorter recruitment time series would be even more vulnerable. Thus, fisheries scientists attempting to develop forecast models should use caution when determining the number of hidden neurons for ANN models.

It is interesting that the SSB covariate was not selected as a significant covariate by a single statistical method considered in this study. This may signal a lack of density dependence in the GOA pollock recruitment process. Although many fisheries management models assume density dependence, it is not often clearly identifiable in marine populations, and the presence of its effect for some species is debatable (Rose et al. 2001). One possible reason for the lack of significant density dependence in GOA pollock recruitment is that the intensity of influential external factors can obscure its relatively weak signal. Whether or not there is a density-dependent effect on GOA pollock recruitment, we still recommend including SSB in recruitment forecast models for two reasons. First, there is a better theoretical and biological basis for doing so. No spawning implies no reproduction. Second, having SSB in the forecast model consistently produced equivalent or better forecasts across the models of same response variable and modeling technique, as seen in our Monte Carlo resampling tests, even though SSB was not a statistically significant covariate.

In building a recruitment forecast model, there are two ways to include SSB: constraining it to be included in the model from the beginning of the variable selection procedure or adding SSB to the model at the final stage after the variable selection procedure completes selecting other covariates. These two approaches may end up having identical or different sets of covariates for the final model. In our study,

for the models with log(recruitment) as the response variable, the sets of selected covariates were different for the same type of statistical methods, depending on whether or not the SSB covariate was constrained. However, the selected covariates were identical (except SSB) for the log(survival) model. Although the modeler is free to choose a variable selection approach, we recommend including the SSB covariate throughout the variable selection procedure for the reason mentioned in the previous paragraph (no spawning, no recruitment) and for the following rationale: SSB's underlying dynamics with other covariates may allow the final model to produce a different set of variables than the model selected without a constraint on SSB. In that case, the final model candidate with SSB included from the very beginning stage of variable selection would be biologically more reasonable than the other model candidate with SSB added at the last stage of variable selection.

It also is noteworthy that no covariate affecting the stock during the spawning season was selected by either MLRs or GAMs. Kim and Gunderson (1989), Bailey et al. (1996), and Kendall et al. (1996) show that GOA pollock recruitment strength is determined mainly during the spawning period. However, Bailey (2000) later reports that the critical stage for the determination of GOA pollock recruitment strength shifted from the larval stage to the juvenile stage after a major environmental regime shift in the late 1970s. Mortality during the larval stage was largely influenced by environmental conditions, and its correlation with recruitment remained good until the mid-1980s. After that, greater juvenile mortality coincided with an increase in the abundance of predatory flatfish species (e.g., arrowtooth flounder (*Atheresthes stomias*) and cod (*Gadus morhua*)). For the postspawning season, the PDO index, which represents long-term variability in the system as an indicator of regime shifts, was selected by our forecast models. A positive value of PDO indicates a warmer regime and a negative value indicates a colder regime in decadal variability of the GOA environmental system. The inverse relationship between PDO and recruitment in MLRs and GAMs suggests that recruitment strength weakens as the ecosystem approaches the warmer regime. Regime shifts alter the dynamics of physical environments and biological systems in major ways (Hollowed and Wooster 1995; Francis et al. 1998; McGowan et al. 1998). The fact that the PDO during postspawning season was selected by our models as an influential covariate affecting recruitment may be further evidence for a shift in the critical period of recruitment strength determination.

Table 4. The means of MAE of forecasting models having log(survival) as response for the resampled out-of-sample data sets. Means are sorted in ascending order. Means that are grouped by underlining are not significantly different at the significance level of 0.05, based on SNK multiple comparison test.

	Models					
	MS	MSC	AS	GS	GSC	ASC
Mean	0.616	0.630	0.646	<u>0.649</u>	0.671	0.672

It may not be feasible to make direct comparisons with Bailey's (2000) results on the shift in the recruitment critical period because this current study uses different modeling approaches and has limited biological variables available. However, both studies indicate that conditions during the postspawning season considerably affect recruitment success. The timing of the during-spawning season includes the larval stage, whereas the postspawning season matches with the juvenile stage. Two postspawning environmental variables were selected as influential by the forecast models in this study: wind mixing and PDO. As no biological variables other than SSB were used in building the forecast models in this study, it is impossible to say that the timing of recruitment determination has shifted to the postspawning season as found in Bailey's study. However, we deduce that wind mixing and PDO might operate indirectly on the dynamics of the pollock population or between the pollock population and other predatory species. Environmental conditions can facilitate retention of juveniles in favorable nursery grounds for better survival or force overlap in habitats of juvenile pollock and predators, leading to favorable conditions for piscivores and thus higher juvenile pollock mortality.

Although the models in this study were not designed to examine recruitment processes and dynamics explicitly, the covariates selected in this study were similar to the covariates identified as influential in other studies. Previous research by Megrey et al. (1995) based on the GOA pollock recruitment time series of 1962–1989 (28 annual data points) found that GOA age-2 pollock recruitment variability was related to precipitation, an index of atmospheric sea-level pressure gradient, and local wind mixing. Our findings, based on an extended time series from 1961–1995 (35 annual recruitment data points were used for model development in this study), are very similar to their results. Bailey et al. (1996) reported that mortality rates of early-feeding larvae tend to be negatively correlated with temperature and positively correlated with wind mixing. Their findings are consistent with our model results in that SST had a negative effect and local wind mixing had a positive effect on recruitment strength during the prespawning season. In our study, these covariates were selected as influential variables by both MLRs and GAMs. Interestingly, the positive effect of wind mixing during prespawning was reversed to a negative effect during the postspawning season in the MLRs and GAMs. It is hypothesized that winds mixing the water column prior to the spawning season establish conditions on the spawning grounds more conducive to larval survival. Wind mixing deepens the mixed layer of nutrient-rich water and consequently enhances primary and secondary production. It has also been shown that strong winds relate to for-

mation and maintenance of eddies in Shelikof Strait, and eddies may play an important role in keeping pollock larvae in favorable nursery locations (Stabeno et al. 1995). Bailey and Macklin (1994) reported that strong wind mixing during pollock larvae's first-feeding period was associated with lower survival, and higher survival rates occurred during calm periods often bracketed by strong mixing. Strong wind mixing during the postspawning season can decrease survival of postlarval and early juvenile stages because young pollock avoid turbulence (Olla and Davis 1990) and avoidance of turbulence can interfere with their ability to feed. Also, strong winds associated with wind mixing can increase ocean circulation causing offshore transport of organisms to unfavorable nursery grounds. These simplistic hypotheses require testing with further studies.

Although there are similarities between this and previous studies in relationships determined among key environmental covariates and recruitment success, care must be taken in their interpretation. In this study, for example, data on biological variables and the geographic extent of environmental variables used for model development were limited. The only biological variable used as a covariate was SSB, and the environmental variables were proxies centered around Shelikof Strait, assuming that recruitment success is largely determined from that area. However, recruitment abundance and spawning biomass data were from the stock assessment model that utilizes data sources from different areas of the GOA, including multiple spawning grounds (Dorn et al. 2003). The Alaska Fisheries Science Center's echo-integration trawl surveys show that the Shelikof area contributes 59% of the total biomass of GOA pollock (Guttorpsen et al. 2003). Thus, relating environmental variables from a confined area, such as Shelikof Strait, to recruitment data that are derived from gulf-wide sources may obscure model results. It is difficult to rule out artifacts in model results due to scale mismatch between recruitment data and covariates (Bailey et al. 2005).

The aim of this study was to build and test recruitment forecast models, not to study recruitment processes and mechanisms, nor to identify key environmental covariates that describe recruitment variability. Thus, it may not be appropriate to draw conclusions on the meaning and detailed interpretation of covariates selected in the models. Rather, the results of covariate selection should be used for generating new hypotheses and for improving forecasting performance through the process of verifying new hypotheses. Even though the forecast models may not help explain functional relationships between response and explanatory variables in a natural ecosystem and the equations may not accurately represent natural processes, it is still useful and advanta-

geous to predict the value of a response variable given the value of associated explanatory variables (Zar 1999). For these reasons, we focus more on the performance of forecasting in selection of covariates for building recruitment forecast models. However, care should be taken to avoid spurious correlations between recruitment and environmental covariates. In a simulation study of recruitment models, Megré et al. (2005) showed that spurious correlations tend to occur more frequently as larger variability is associated with recruitment and environmental covariates.

We expected that unconventional computer intensive methods, such as GAMs and ANNs, would be better for recruitment forecasting than conventional parametric methods, because GAMs and ANNs have comparatively relaxed statistical assumptions and flexibility in function approximation. Contrary to our expectations, conventional parametric MLR statistical models outperformed more sophisticated GAMs and ANNs in forecasting recruitment strength when compared using a Monte Carlo resampling method. The apparent superiority of MLRs in recruitment forecasting may be limited to the GOA pollock case for the available data and should not be generalized to all other fish species, because recruitment characteristics are species-, spatial-, and temporal-specific. Different nonparametric techniques might prove to be better forecasting tools for other species. However, use of Monte Carlo resampling for evaluation of possible recruitment forecasting model candidates can be useful and effective if applied to other fish species and should be applied when determining the best forecasting model. This type of resampling approach can also be used to gauge the robustness of the models through cross-validation and to identify key covariates through sensitivity analysis (Lunneborg 2000).

We advocate employing more than one method to build recruitment forecast models. Each statistical model has advantages and disadvantages. As Prager (2003) points out, each model provides, from its particular perspective, an imperfect view of reality. The more perspectives one gains, the better. If the merits of multiple modeling techniques could be combined, we would be able to build better-performing recruitment forecast models instead of relying on just one modeling technique. For example, GAMs are flexible and suggest the functional relationship of contributing covariates, but they do not indicate the influence of the covariate to the overall prediction. MLRs are straightforward and reliable for parameter estimation and hypothesis testing, but they require rather restrictive modeling assumptions, such as linearity, normality, and homogeneous variances. Thus, a modeler can use GAMs to explore functional relationships and MLRs to test for the significance of covariates' coefficients in quadratic or cubic terms if some covariates appear to be nonlinearly related to recruitment in GAMs. As demonstrated in this study, once prospective model candidates are identified, a Monte Carlo resampling strategy can be applied to evaluate the models' performance for forecasting.

Acknowledgements

This publication was funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO), University of Washington, under NOAA Cooperative Agreement No. NA17RJ1232, Contribution No. 1417. This is contri-

bution No. EcoFOCI-L610 to NOAA's Ecosystems and Fisheries – Oceanography Coordinated Investigations and contribution No. 2968 to the Pacific Marine Environmental Laboratory. Y.-W. Lee was funded by the NOAA Steller Sea Lion Research Initiative. We thank Mick Spillane at JISAO for assembling the environmental data for analysis. We also thank two anonymous reviewers and an associate editor, Stephen Smith, for their constructive review comments. Y.-W. Lee thanks Wan-Ying Chang at the Washington Department of Fish and Wildlife for the insightful discussions on the statistics during the review processes.

References

- Akaike, H. 1981. Likelihood of a model and information criteria. *J. Econom.* **16**: 3–14. doi:10.1016/0304-4076(81)90071-3.
- Akoi, I., and Komatsu, T. 1997. Analysis and prediction of the fluctuation of sardine abundance using a neural network. *Oceanol. Acta*, **20**: 81–88.
- Bailey, K.M. 1989. Interaction between the vertical distribution of juvenile walleye pollock *Theragra chalcogramma* in the eastern Bering Sea, and cannibalism. *Mar. Ecol. Prog. Ser.* **53**: 205–213. doi:10.3354/meps053205.
- Bailey, K.M. 2000. Shifting control of recruitment of walleye pollock *Theragra chalcogramma* after a major climatic and ecosystem change. *Mar. Ecol. Prog. Ser.* **189**: 117–123.
- Bailey, K.M., and Macklin, S.A. 1994. Analysis of patterns in larval walleye pollock *Theragra chalcogramma* survival and wind mixing events in Shelikof Strait, Gulf of Alaska. *Mar. Ecol. Prog. Ser.* **113**: 1–12. doi:10.3354/meps113001.
- Bailey, K.M., Picquelle, S., and Spring, S. 1996. Mortality of larval walleye pollock *Theragra chalcogramma* in the western Gulf of Alaska. *Fish. Oceanogr.* **5**(Suppl. 1): 124–136. doi:10.1111/j.1365-2419.1996.tb00087.x.
- Bailey, K.M., Ciannelli, L., Bond, N.A., Belgrano, A., and Stenseth, N.C. 2005. Recruitment of walleye pollock in a physically and biologically complex ecosystem: a new perspective. *Prog. Oceanogr.* **67**: 24–42. doi:10.1016/j.pocean.2005.06.001.
- Beare, D.J., and Reid, D.G. 2002. Investigating spatio-temporal change in spawning activity by Atlantic mackerel between 1977 and 1998 using Generalized Additive Models. *ICES J. Mar. Sci.* **59**: 711–724. doi:10.1006/jmsc.2002.1207.
- Beverton, R.J.H., and Holt, S.J. 1957. On the dynamics of exploited fish populations. *Fisheries Investigations. Ser. 2. Vol. 19*. UK Ministry of Agriculture and Fisheries, Chapman and Hall, London, UK.
- Brockwell, P.J., and Davis, R.A. 2003. *Introduction to time series and forecasting*. 2nd ed. Springer Inc., New York.
- Burnham, K.P., and Anderson, D.R. 2002. *Model selection and multi-model inference*. Springer Inc., New York.
- Cardinale, M., and Arrhenius, F. 2000. The influence of stock structure and environmental conditions on the recruitment process of Baltic cod estimated using a generalized additive model. *Can. J. Fish. Aquat. Sci.* **57**: 2402–2409. doi:10.1139/cjfas-57-12-2402.
- Chambers, J.M., and Hastie, T.J. (Editors). 1992. *Statistical models in S*. Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, Calif.
- Chen, D.G., and Ware, D.M. 1999. A neural network model for forecasting fish stock recruitment. *Can. J. Fish. Aquat. Sci.* **56**: 2385–2396. doi:10.1139/cjfas-56-12-2385.
- Ciannelli, L., Bailey, K.M., Chan, K.S., and Stenseth, N.C. 2007. Phenological and geographical patterns of walleye pollock (*Ther-*

- agra chalcogramma*) spawning in the western Gulf of Alaska. Can. J. Fish. Aquat. Sci. **64**: 713–722. doi:10.1139/F07-049.
- Cushing, D.H. 1988. The study of stock and recruitment. In *Fish population dynamics*. 2nd ed. Edited by J.A. Gulland. John Wiley and Sons Ltd., Chichester, UK. pp. 105–128.
- Daskalov, G. 1999. Relating fish recruitment to stock biomass and physical environment in the Black Sea using generalized additive models. Fish. Res. **41**: 1–23. doi:10.1016/S0165-7836(99)00006-5.
- Dorn, M., Barbeaux, S., Guttormsen, M., Megrey, B.A., Hollowed, A., Brown, E., and Spalinger, K. 2003. Assessment of walleye pollock in the Gulf of Alaska. In *Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska*, 2003. North Pacific Fisheries Management Council, Anchorage, Alaska. pp. 33–124.
- Draper, N.R., and Smith, H. 1998. Applied regression analysis. 3rd ed. Wiley, New York.
- Emery, W.J., and Hamilton, K. 1985. Atmospheric forcing of interannual variability in the Northeast Pacific Ocean: connections with El Niño. J. Geophys. Res. **90**: 857–868. doi:10.1029/JC090iC01p00857.
- Francis, R.C., Hare, S.R., Hollowed, A.B., and Wooster, W.S. 1998. Effects of interdecadal climate variability on the oceanic ecosystems of the Northeast Pacific. Fish. Oceanogr. **7**: 1–21. doi:10.1046/j.1365-2419.1998.00052.x.
- Guttormsen, M.A., Wilson, C.D., and Stienessen, S. 2003. Results of the February and March 2003 echo integration-trawl surveys of walleye pollock (*Theragra chalcogramma*) conducted in the Gulf of Alaska, Cruises MF2003-01 and MF2003-05. In *Stock Assessment and Fishery Evaluation Report for Gulf of Alaska*. Prepared by the Gulf of Alaska Groundfish Plan Team, North Pacific Fishery Management Council, P.O. Box 103136, Anchorage, AK 99510. North Pacific Fisheries Management Council, Anchorage, Alaska. pp. 799–844.
- Hastie, T.J., and Tibshirani, R.J. 1990. Generalized additive models. Chapman and Hall, New York.
- Hilborn, R., and Walters, C.J. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Chapman and Hall, New York.
- Hollowed, A.B., and Wooster, W.S. 1995. Decadal-scale variations in the eastern subarctic Pacific. II. Response of Northeast Pacific fish stocks. In *Climate change and northern fish populations*. Edited by R.J. Beamish. Can. Spec. Publ. Fish. Aquat. Sci. No. 121. pp. 373–385.
- Huse, G., and Gjøsæter, H. 1999. A neural network approach for predicting stock abundance of the Barents Sea capelin. Sarsia, **84**: 457–464.
- Huse, G., and Ottersen, G. 2003. Forecasting recruitment and stock biomass of Northeast Arctic cod using neural networks. Sci. Mar. **67**(Suppl. 1): 325–335.
- Jacobson, L.D., and MacCall, A.D. 1995. Stock-recruitment models for Pacific sardine (*Sardinops sagax*). Can. J. Fish. Aquat. Sci. **52**: 566–577. doi:10.1139/f95-057.
- Kendall, A.W., Jr., and Duker, G.J. 1998. The development of recruitment fisheries oceanography in the United States. Fish. Oceanogr. **7**(2): 69–88. doi:10.1046/j.1365-2419.1998.00056.x.
- Kendall, A.W., Jr., Schumacher, J.D., and Kim, S. 1996. Walleye pollock recruitment in Shelikof Strait: applied fisheries oceanography. Fish. Oceanogr. **5**(Suppl. 1): 4–18. doi:10.1111/j.1365-2419.1996.tb00079.x.
- Kim, S., and Gunderson, D.R. 1989. Cohort dynamics of walleye pollock in Shelikof Strait, Gulf of Alaska, during the egg and larval period. Trans. Am. Fish. Soc. **118**: 264–273. doi:10.1577/1548-8659(1989)118<0264:CDOWPI>2.3.CO;2.
- Laë, R., Lek, S., and Moreau, J. 1999. Predicting fish yield of African lakes using neural networks. Ecol. Model. **120**: 325–335. doi:10.1016/S0304-3800(99)00112-X.
- Lunneborg, C.E. 2000. Data analysis by resampling concepts and applications. Duxbury, Pacific Grove, Calif.
- Mallows, C.L. 1973. Some comments on C_p . Technometrics, **12**: 591–612.
- Mantua, N.J., and Hare, S.R. 2002. The Pacific Decadal Oscillation. J. Oceanogr. **58**(1): 35–44. doi:10.1023/A:1015820616384.
- McGowan, J.A., Cayan, D.R., and Dorman, L.M. 1998. Climate-ocean variability and ecosystem response in the Northeast Pacific. Science (Washington, D.C.), **281**: 210–217. doi:10.1126/science.281.5374.210. PMID:9660743.
- Megrey, B.A., Bograd, S.J., Rugen, W.C., Hollowed, A.B., Stabeno, P.J., Macklin, S.A., Schumacher, J.D., and Ingraham, W.J., Jr. 1995. An exploratory analysis of associations between biotic and abiotic factors and year-class strength of Gulf of Alaska walleye pollock. In *Climate change and northern fish populations*. Edited by R.J. Beamish. Can. Spec. Publ. Fish. Aquat. Sci. No. 121. pp. 227–243.
- Megrey, B.A., Hollowed, A.B., Hare, S.R., Macklin, S.A., and Stabeno, P.J. 1996. Contributions of FOCI research to forecasts of year-class strength of walleye pollock in Shelikof Strait, Alaska. Fish. Oceanogr. **5**: 189–203. doi:10.1111/j.1365-2419.1996.tb00092.x.
- Megrey, B.A., Lee, Y.-W., and Macklin, S.A. 2005. Comparative analysis of statistical tools to identify recruitment-environment relationships and forecast recruitment strength. ICES J. Mar. Sci. **62**: 1256–1269. doi:10.1016/j.icesjms.2005.05.018.
- Meuter, F.J., Ladd, C., Palmer, M.C., and Norcross, B.L. 2006. Bottom-up and top-down controls of walleye pollock (*Theragra chalcogramma*) on the eastern Bering Sea shelf. Prog. Oceanogr. **68**: 152–183.
- Myers, R.A., Bridson, J., and Barrowman, H.J. 1995. Summary of worldwide stock and recruitment data. Can. Tech. Rep. Fish. Aquat. Sci. No. 2024.
- Needle, C.L. 2001. Recruitment models: diagnosis and prognosis. Rev. Fish Biol. Fish. **11**: 95–111. doi:10.1023/A:1015208017674.
- Olla, B.L., and Davis, M.W. 1990. Effects of physical factors on the vertical distribution of larval walleye pollock *Theragra chalcogramma* under controlled laboratory conditions. Mar. Ecol. Prog. Ser. **63**: 105–112. doi:10.3354/meps063105.
- Porter, S.M., Ciannelli, L., Hillgruber, N., Bailey, K.M., Chan, K.S., Canino, M.F., and Haldorson, L.J. 2005. Environmental factors influencing larval walleye pollock *Theragra chalcogramma* feeding in Alaskan waters. Mar. Ecol. Prog. Ser. **302**: 207–217. doi:10.3354/meps302207.
- Prager, M.H. 2003. Reply to the letter to the editor. Fish. Res. **61**: 151–154. doi:10.1016/S0165-7836(02)00274-6.
- Rawlings, J.O., Pantula, S.G., and Dickey, D.A. 2005. Applied regression analysis: a research tool. 2nd ed. Springer, Inc., New York.
- Ricker, W.E. 1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Board Can. No. 191.
- Ripley, B.D. 1996. Pattern recognition and neural networks. Cambridge University Press, Cambridge, UK.
- Rose, K.A., Cowan, J.H., Winemiller, K.O., Myers, R.M., and Hilborn, R. 2001. Compensatory density dependence in fish populations: importance, controversy, understanding and prognosis. Fish Fish. **2**: 293–327.
- Rosenkranz, G., Tyler, A.V., and Kruse, G.H. 2001. Effects of water temperature and wind on year-class success of Tanner

- crabs in Bristol Bay, Alaska. Fish. Oceanogr. **10**: 1–12. doi:10.1046/j.1365-2419.2001.00153.x.
- Royer, T.C. 1982. Coastal fresh water discharge in the Northeast Pacific. J. Geophys. Res. **87**: 2017–2021. doi:10.1029/JC087iC03p02017.
- Sarle, W.S. 1994. Neural networks and statistical models. In Proceeding of the Nineteenth Annual SAS Users Group International Conference, 10–13 April 1994, Dallas, Texas. SAS Institute Inc., Cary, N.C. pp. 1–13.
- Schnute, J. 1985. A general theory for analysis of catch and effort data. Can. J. Fish. Aquat. Sci. **42**: 414–429. doi:10.1139/f85-057.
- Schwarz, G. 1978. Estimating the dimension of a model. Ann. Stat. **6**: 461–464. doi:10.1214/aos/1176344136.
- Stabeno, P.J., Hermann, A.J., Bond, N.A., and Bograd, S.J. 1995. Modeling the impact of climate variability on the advection of larval walleye pollock (*Theragra chalcogramma*) in the Gulf of Alaska. In Climate change and northern fish populations. Edited by R.J. Beamish. Can. Spec. Publ. Fish. Aquat. Sci. No. 121. pp. 719–727.
- Swartzman, G., Silverman, E., and Williamson, N. 1995. Relating trends in walleye pollock (*Theragra chalcogramma*) abundance in the Bering Sea to environmental factors. Can. J. Fish. Aquat. Sci. **52**: 369–380. doi:10.1139/f95-039.
- Trenberth, K.E. 1984. Signal versus noise in the Southern Oscillation. Mon. Weather Rev. **112**: 326–332. doi:10.1175/1520-0493(1984)112<0326:SVNITS>2.0.CO;2.
- Venables, W.N., and Ripley, B.D. 2002. Modern applied statistics with S. 4th ed. Springer Inc., New York.
- Walters, C.J., and Ludwig, D. 1981. Effects of measurements errors on the assessment of stock-recruitment relationships. Can. J. Fish. Aquat. Sci. **38**: 704–710. doi:10.1139/f81-093.
- Wespestad, V.G., Fritz, L.W., Ingraham, W.J., and Megrey, B.A. 2000. On relationships between cannibalism, climate variability, physical transport, and recruitment success of Bering Sea walleye pollock (*Theragra chalcogramma*). ICES J. Mar. Sci. **57**: 272–278. doi:10.1006/jmsc.2000.0640.
- Zar, J.H. 1999. Biostatistical analysis. 4th ed. Prentice-Hall, Inc., Englewood Cliffs, N.J.