

Is it possible to predict habitat use by spawning salmonids? A test using California golden trout (*Oncorhynchus mykiss aguabonita*)

Roland A. Knapp and Haiganoush K. Preisler

Abstract: It is widely believed that stream salmonids select spawning sites based on water depth, water velocity, and substrate size. Attempts to predict spawning locations using these habitat features have met with little success, however. In this study, we used nonparametric logistic regression to determine what habitat features were associated with the locations chosen by spawning California golden trout (*Oncorhynchus mykiss aguabonita*). From this nonparametric model, we developed a parametric model that incorporated the habitat features most strongly associated with spawning sites and used this model to calculate the probability of use by spawning golden trout for specific stream locations. The overall nonparametric model was highly significant and explained 62% of the variation in spawning location. Four of the eight habitat variables, substrate size, water depth, water velocity, and stream width, had highly significant effects and alone explained 59% of the variation in spawning location. The results of a cross-validation procedure indicated that the parametric model generally provided a good fit to the data. These results indicate that location-specific probabilities of use were predictable based on easily measured habitat characteristics and that nonparametric regression, an approach still rarely used in ecological studies, may have considerable utility in the development of fish-habitat models. Given the escalating pace at which fish habitats are being altered, such models are increasingly important in predicting the effects of these alterations on populations.

Résumé : Il semble bien établi que les salmonidés des cours d'eau choisissent leur frayère selon la profondeur de l'eau, la vitesse du courant et la taille du substrat. Cependant, des essais de prédiction de l'emplacement des frayères en utilisant ces caractéristiques de l'habitat ont échoué. La présente étude utilise la régression logistique non paramétrique pour déterminer les caractéristiques de l'habitat associées aux frayères choisies par la truite dorée (*Oncorhynchus mykiss aguabonita*) de Californie. À l'aide du modèle non paramétrique, nous avons développé un modèle paramétrique qui tient compte des caractéristiques de l'habitat les plus étroitement associées aux lieux de fraye et nous l'avons utilisé pour calculer la probabilité d'utilisation par les truites dorées pour la fraye d'endroits spécifiques dans les cours d'eau. En général, le modèle non paramétrique était très représentatif et expliquait 62 % des variations dans le choix des frayères. Quatre des huit variables (taille du substrat, profondeur de l'eau, vitesse du courant et largeur du cours d'eau) avaient des effets extrêmement importants et, à elles seules, expliquaient 59 % des variations dans le choix des frayères. Les résultats d'une procédure de contrevalidation ont montré que le modèle paramétrique permettait d'ajuster les données de façon satisfaisante. Ces résultats ont aussi fait ressortir que la probabilité d'utilisation d'un lieu spécifique pouvait être prédite en fonction de caractéristiques de l'habitat facilement mesurables et que la régression non paramétrique, approche encore peu utilisée dans les études écologiques, pouvait être très utile pour développer des modèles d'habitat du poisson. Le rythme accéléré auquel les habitats du poisson sont modifiés rend indispensables des modèles qui prévoient les effets de ces modifications sur les populations.

[Traduit par la Rédaction]

Introduction

The effects of habitat conditions on stream fish populations are frequently evaluated using microhabitat models (e.g., Instream Flow Incremental Methodology (IFIM) and associated Physical Habitat Simulation (PHABSIM) System; Bovee 1978; Reiser et al. 1989). These models are based on

the assumption that fish selectively occupy habitats as a function of their preferences for habitat characteristics such as water depth, water velocity, and substrate type (Stalnaker 1979); if true, then by knowing the distributions of these habitat features and the preferences of fish for these features, the distribution of fish within a stream should be predictable (Beecher et al. 1993). While several studies based on fish

Received April 27, 1998. Accepted May 14, 1999.
J14558

R.A. Knapp.¹ Sierra Nevada Aquatic Research Laboratory, University of California, Star Route 1, Box 198, Mammoth Lakes, CA 93546, U.S.A., and the Marine Science Institute, University of California, Santa Barbara, CA 93106, U.S.A.

H.K. Preisler. Pacific Southwest Research Station, USDA Forest Service, Box 245, Berkeley, CA 94701-0245, U.S.A.

¹Author for all correspondence at the following address: Sierra Nevada Aquatic Research Laboratory, University of California, Star Route 1, Box 198, Mammoth Lakes, CA 93546, U.S.A.

distributions measured during the nonreproductive season have provided support for this central tenet (Beecher et al. 1993; Nehring and Anderson 1993), there have been very few attempts to predict locations chosen by spawning fish. In the only such published study that we have found, Shirvell (1989) reported a very poor correspondence between the locations predicted by measurements of water depth, water velocity, and substrate size as containing suitable spawning habitat and the locations actually used by spawning fish. This poor model performance has led some to suggest that stream salmonids select spawning locations based on habitat characteristics other than or in addition to depth, velocity, and substrate (Shirvell 1989; Beecher et al. 1993). Indeed, several studies have indicated the importance of groundwater seepage (Witzel and MacCrimmon 1983; Lorenz and Eiler 1989), river bottom topography (Shirvell 1989), and proximity to streambanks (Stefferd 1993) in influencing spawning site selection. Despite the uncertainty over the relative importance of these habitat characteristics in influencing spawning locations, important stream management decisions such as instream flow recommendations are routinely made based in part on spawning habitat availability as predicted solely by water depth, water velocity, and substrate size (e.g., California Department of Fish and Game 1996).

The goal of the present study was to first determine what habitat characteristics influence the locations chosen by spawning California golden trout (*Oncorhynchus mykiss aguabonita*) and then test the hypothesis that the probability of individual stream locations being used by spawning fish is predictable based on these fish-habitat relationships. To test this hypothesis, we measured channel and microhabitat characteristics in a stream located in the southern Sierra Nevada, California, and then used nonparametric logistic regression to identify those stream features associated with the presence of golden trout redds. After identifying which of the measured variables were most closely associated with spawning locations, we developed a parametric model based on these variables to predict the probability of use by spawning golden trout. We then compared these predicted probabilities of use with the actual locations used by spawning golden trout.

Materials and methods

Study site

Mulkey Creek is a tributary to the upper South Fork Kern River, a watershed located at the southern end of the Sierra Nevada, Inyo National Forest, California (118°15'N, 36°22'W). The portion of Mulkey Creek used in this study was located at the lower end of Mulkey Meadow at an elevation of about 2850 m. A map and detailed description of the study area are provided in Knapp and Vredenburg (1996) and Knapp et al. (1998). The California golden trout is the only fish species present in Mulkey Creek.

Two sections of Mulkey Creek were used in this study and were chosen because of their similarly low stream gradients, gradients typical of the meadow streams in the South Fork Kern River drainage in which the golden trout is native. Each of these study sections was about 750 m long and, based on the frequent observation of spawning activity in these sections during previous years (Knapp and Vredenburg 1996), contained abundant spawning habitat. These two study sections were separated by a 1.5-km stream

section characterized by a substantially higher stream gradient. Because high stream gradients are unusual in the meadow streams of the South Fork Kern River drainage, we did not use this section in our study. In the two study sections, the stream channel alternated between narrow, deep reaches (E channels; Rosgen 1994) and wide, shallow reaches (C channels; Rosgen 1994). Based on these channel types, we divided the stream within the two study sections into 11 distinct reaches (six C channel reaches and five E channel reaches). Reaches ranged in length from 45 to 130 m. We made measurements of channel morphology and stream microhabitat characteristics in all 11 reaches and used these data to develop the logistic regression model.

Data collection

We measured characteristics of redds and stream channels immediately after the 1996 spawning period to ensure that measurements reflected those encountered by spawning trout (end of spawning: June 5; habitat measurements: June 6–12; Knapp et al. 1998). All channel and instream characteristics were measured along transects spaced 2.5 m apart and arranged perpendicular to stream flow (number of transects per reach = 20–46). At each transect, we measured stream width, bank angle, and streambank vegetation cover. Stream width (excluding islands) was measured to the nearest 5 cm with a meter tape. We measured the angle of both banks to the nearest 5° using a clinometer on a 1.5-m rod placed against the bank slope (Platts et al. 1983); lower angles indicate more bank overhang. Bank vegetation cover was estimated visually on both banks inside a 0.1-m² area extending from the stream edge to 50 cm away from the stream and 10 cm upstream and downstream of the transect.

We measured water depth (to the nearest 2 cm), water velocity, and substrate size at five equally spaced points along each transect. Water velocity was measured with an electromagnetic current meter to the nearest 1 cm·s⁻¹ at 60% of the water depth, and each measurement represented a 10-s average. We visually estimated the average substrate size within a 10 × 10 cm area centered at each point. Visual estimates of average substrate size were highly correlated with the actual geometric mean diameter calculated from core samples ($n = 20$, $R^2 = 0.93$, $P < 0.0001$; Knapp et al. 1998).

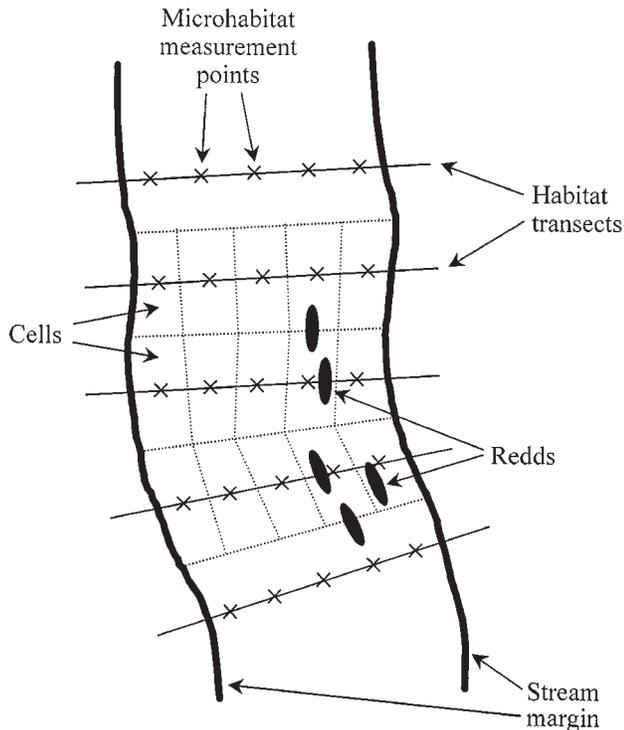
Stream channel maps for the study reaches were made by tracing the stream margins from high-resolution aerial photographs of the study area. Locations of transects, microhabitat measurement points, and redds were drawn onto these maps (Fig. 1) and the maps were then used to divide the study reaches into a series of "cells." A cell was a roughly rectangular portion of the stream surface whose dimensions were determined by the spacing of the stream channel transects and the five microhabitat measurement points across each transect (Fig. 1). Each cell had at its center one of the microhabitat measurement points and extended on average 1.25 m upstream (i.e., half way to the next transect) and 1.25 m downstream and was one fifth as wide as the stream at that transect (Fig. 1). Water depth, velocity, and substrate size from each microhabitat measurement point were assigned to the associated cell. A total of 1765 cells were used in this study, with individual study reaches containing 100–230 cells. We considered a cell as having been utilized by spawning fish if it contained any portion of a redd. The position of each redd relative to the streambank was estimated as the distance from the center of the cell containing the redd to the nearest streambank.

Statistical analyses

Habitat description

Differences in habitat characteristics associated with cells used and not used by spawning golden trout were analyzed using the nonparametric Kruskal–Wallis one-way analysis of variance. We used a nonparametric test because habitat variables were typically

Fig. 1. Portion of a stream channel map showing how cells were created from habitat transects and microhabitat measurement points.



not normally distributed and had unequal variances, and normality and variance equivalency generally could not be achieved using standard transformations.

Model development

We used nonparametric logistic regression to describe the relationship between the independent variables (channel and microhabitat characteristics) and the probability of finding redds at a given location (probability of response). Nonparametric logistic regression models are a subclass of generalized additive models (GAMs) and as such relax the assumptions of ordinary linear regression in two ways: (i) the distribution of the dependent variable need not be symmetric or homoscedastic and (ii) the relationship between the dependent and independent variables need not be linear. Relaxation of the assumption of a linear relationship between the dependent and independent variables is accomplished using a spline smoothing function (e.g., loess) to determine the fitted model that best fits the independent variables (Cleveland and Devlin 1988; Efron and Tibshirani 1991). Nonparametric logistic regression approaches are likely to be of considerable utility in studies of habitat selection by fishes, since numerous response-curve shapes are typically plausible but the actual shape will likely be unknown prior to analysis. Researchers have only recently begun to use nonparametric regression models in ecological studies, however (e.g., Wiser et al. 1998).

Because the shapes of the curves describing the probability of response in our study were unknown prior to analysis and we had no reason to believe that the relationships were linear, nonparametric logistic regression was ideally suited for analyzing the effect of habitat characteristics on the probability of finding redds at a given location. Specifically, if p_i is the probability of finding redds at location i , then

$$(1) \quad p_i = \frac{e^{\theta_i}}{1 + e^{\theta_i}}$$

where the linear predictor (i.e., logit line) θ_i is some function of the independent variables. The relationship we used for θ_i was as follows:

$$(2) \quad \theta_i = \beta_0 + g_1(\text{water depth}_i) + g_2(\text{water velocity}_i) \\ + g_3(\text{substrate size}_i) + g_4(\text{stream width}_i) \\ + g_5(\text{bank angle}_i) + g_6(\text{bank vegetation cover}_i) \\ + g_7(\text{distance to bank}_i) + g_8(\text{cell location}_i)$$

where $g(\cdot)$ is a nonparametric smooth function that characterizes the effect of an independent variable on the probability of response (Venables and Ripley 1994). The nonparametric functions for each of the independent variables were fitted using a loess smoothing function.

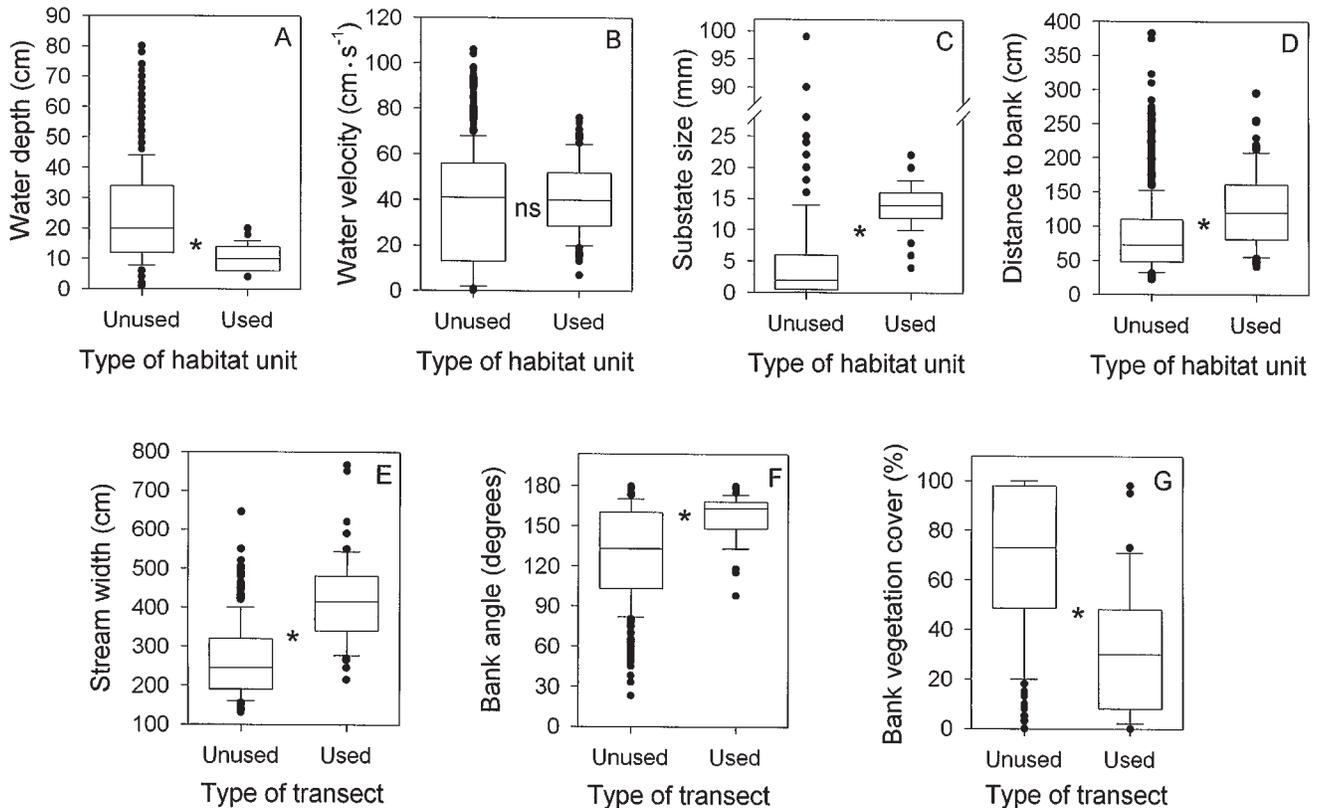
Spatial autocorrelation is a common statistical property of ecological variables and presents a problem for statistical testing because autocorrelated data violate the assumption of independence common to most standard statistical procedures (Legendre 1993; Thomson et al. 1996). We expected spatial autocorrelation within our study reaches because neighboring cells are likely to be more similar in habitat characteristics than are cells separated by a greater distance. One way to overcome this problem is to explicitly incorporate spatial structure into statistical models (Legendre 1993). Based on the approach used by Hobert et al. (1997), we incorporated spatial dependence by including a nonlinear spatial term ("cell location") on equal footing with the other independent variables. We defined cell location as the distance from the downstream-most transect of reach 1 (the downstream-most study reach) to the habitat transect containing the cell of interest. If spatial structure is important, the inclusion of the variable "cell location" would be expected to significantly increase the amount of variation explained by the model.

The best combination of independent variables was determined by evaluating the change in deviance (McCullagh and Nelder 1989) resulting from dropping each variable from the model in the presence of all other variables. Analysis of deviance and likelihood ratio tests (based on the binomial distribution) were used to test the significance of each of the independent variables on the probability of redd presence (McCullagh and Nelder 1989). The relative importance of significant variables was determined by calculating the Akaike information criteria (AIC; Linhart and Zucchini 1986). Larger AIC values indicate a greater relative importance. In order to evaluate the explanatory power provided by each independent variable, we also calculated the coefficient of determination (R^2) for each independent variable (Nagelkerke 1991). Values of R^2 were calculated by adding variables to the model in the order of their associated AIC value, such that the independent variable with the largest AIC value was added first and the variable with the smallest AIC value was added last.

Although nonparametric logistic regression is useful for evaluating the relative importance of independent variables and for describing the shape of response curves, parametric models are preferable in developing a predictive model if they provide an adequate fit to the data (Yee and Mitchell 1991). To develop a parametric model from the nonparametric model, we first examined plots of the response curves generated by the nonparametric regression for the four highly significant independent variables ($P < 0.0001$). Based on the shapes of these response curves, we determined the appropriate parametric model (e.g., linear, polynomial, logarithmic) to approximate the nonparametric functions. We based the parametric model solely on the highly significant independent variables to ensure that only those independent variables with high explanatory power were included in the model.

To assess whether the parametric model predicted the probability of use (p_i) as adequately as the nonparametric model (i.e., that

Fig. 2. Box plots showing (A) water depth, (B) water velocity, (C) substrate size, (D) distance to the nearest streambank, (E) stream width, (F) bank angle, and (G) bank vegetation cover for cells or transects that were unused versus used by spawning golden trout. The bottom and top of each box indicate the 25th and 75th percentiles, respectively, the line within each box marks the median, and the whiskers below and above each box indicate the 10th and 90th percentiles, respectively. Observations falling outside these percentiles are shown as points above and below the whiskers. An asterisk between boxes indicates that the difference between unused and used cells was highly significant ($P < 0.0001$), and “ns” indicates that the difference was not significant ($P > 0.05$). Significance was determined using a Kruskal–Wallis one-way analysis of variance. Data are from all 11 study reaches combined. For water depth, water velocity, substrate size, and distance to the nearest streambank, $n_{\text{unused}} = 1674$ and $n_{\text{used}} = 91$. For stream width, bank angle, and bank vegetation cover, $n_{\text{unused}} = 291$ and $n_{\text{used}} = 62$.



the functional forms that we used in the parametric model accurately represented the response curves generated by the non-parametric regression), the total deviances of the two models were compared. Since the fit of the two models was similar, we then used the parametric model to calculate, for each cell, the probability of use by spawning golden trout. All regression-related calculations were done using the S-Plus statistical package (MathSoft 1997).

Model cross-validation

To determine how well the parametric model fit data from stream reaches not used in model development, we used a cross-validation procedure (Breiman et al. 1984) followed by χ^2 goodness-of-fit tests (Hosmer and Lemeshow 1989). We first dropped one of the 11 study reaches from the data set and estimated model parameters using the remaining 10 reaches. We then used this reduced model to calculate the p_i values for all cells in the dropped reach. This procedure was repeated 10 times, each time dropping a different study reach from the model. We then divided all cells from each of the dropped reaches into 22 groups based on their predicted p_i values. The goodness-of-fit between the observed and predicted probabilities of redd presence for each of these 22 groups was determined using the Pearson χ^2 statistic, which was then compared with a χ^2 distribution with 14 degrees of freedom (degrees of freedom = number of groups – number of parameters esti-

mated in model = 22 – 8). A small P value indicates that the model does not provide a good fit to the data.

Visualization of model fit

To visualize the fit of the parametric model, we selected one of the 11 study reaches and compared the probability of use associated with each cell in this study reach with the locations actually used by spawning golden trout. For this analysis, we purposefully selected the study reach to which the model provided the poorest fit (reach 9), since this allowed a more thorough evaluation of the reasons underlying any discrepancies between p_i values and actual redd locations than if we had used a study reach to which the model provided a better fit.

Results

Golden trout spawning activity began on May 28 and only a few females were still observed on redds by June 5. No spawning was seen after June 8. Females constructed a total of 153 redds in the study reaches (0–67 per reach). Since each cell could contain more than one redd, the number of redd-containing cells in the study reaches was 91 (5%).

Average habitat values for cells used and not used by golden trout were significantly different for all measured

Table 1. Analysis of deviance table showing the statistical significance (P value), AIC, and R^2 values of the independent variables in the nonparametric logistic regression model.

Model	Model		Test		P value	AIC ^c	R^2 ^d
	Deviance	df	Deviance ^a	df ^b			
Null model	717	1764					
Full model	308	1747.6					62.0
Full model less:							
substrate size	457	1750.0	149	2.5	$<10^{-10}$	487.0	31.6
water depth	359	1748.6	51	1.0	$<10^{-10}$	391.8	49.4
water velocity	334	1749.2	26	1.6	9.5×10^{-7}	365.6	55.2
stream width	329	1749.5	21	1.9	2.5×10^{-5}	360.0	59.1
cell location	319	1751.3	11	3.7	0.02	346.4	60.3
vegetation cover	315	1749.4	7	1.8	0.02	346.2	61.2
bank angle	311	1749.2	3	1.6	0.11	342.6	61.6
distance to bank	310	1749.5	2	1.9	0.22	341.0	62.0

^aTest deviance = (deviance of full model less one covariate) – (deviance of full model).

^bTest df = (df of full model less one covariate) – (df of full model).

^cAIC = (deviance of full model less one covariate) + 2(1765 – df of full model less one covariate).

^dThe R^2 value shown for each variable represents the R^2 of a model that includes the variable and all other variables with higher AIC scores.

variables except water velocity (Fig. 2). Cells used for spawning had on average significantly shallower water depths and larger substrates and were further from stream-banks than unused cells (Figs. 2A, 2C, and 2D). In addition, transects containing at least one cell used for spawning were characterized by larger stream widths, larger bank angles (i.e., less bank overhang), and less bank vegetation cover than transects containing only unused cells (Figs. 2E, 2F, and 2G).

The overall nonparametric logistic regression model (eq. 2) was highly significant ($P < 10^{-10}$) and explained 62% of the variation in redd location (Table 1). The independent variables water depth, water velocity, substrate size, and stream width all had highly significant effects ($P < 0.0001$) on the probability of redd presence, while cell location and bank vegetation cover had marginally significant effects (Table 1). Bank angle and distance to bank did not contribute significantly to the model ($P > 0.05$) (Table 1). Based on the AIC (Table 1), the relative importance of the significant variables in explaining the location of redds was substrate size > water depth > water velocity > stream width > cell location > bank vegetation cover. The four highly significant independent variables alone accounted for 59% of the variation in redd location, while the two marginally significant independent variables, cell location and bank vegetation cover, together accounted for only an additional 2% of the variation (Table 1).

The relationships between p_i and the highly significant independent variables all had substantial nonlinear contributions (Fig. 3). The response shapes indicated that the probability of redd presence was highest in areas with intermediate substrate sizes (Fig. 3A), the shallowest water depths (Fig. 3B), intermediate water velocities (Fig. 3C), and the largest stream widths (Fig. 3D). We approximated these response shapes with second-degree polynomials (for substrate size, water depth, and water velocity) and a logarithmic function (for stream width). The difference in the deviance between the nonparametric and parametric regression models was not significant ($P > 0.05$), indicating that our

choice of functional forms closely approximated the actual form of each significant relationship. The parametric model, based on water depth, velocity, substrate size, and stream width, is represented by the following equation:

$$(3) \hat{\theta} = -29.44 + 0.63 (\text{water depth}) - 0.03 (\text{water depth})^2 \\ + 0.16 (\text{water velocity}) - 0.002 (\text{water velocity})^2 \\ + 1.52 (\text{substrate size}) - 0.05 (\text{substrate size})^2 \\ + 2.17 (\log_e [\text{stream width}]).$$

The results of the cross-validation procedure indicated that the parametric model generally provided a good fit to study reaches not used for model parameterization. For the 22 categories of the linear predictor, the observed and predicted probabilities of use were very similar, and almost all observed values were within the 95% confidence interval (Fig. 4). Despite the reasonably close match between observed and predicted probabilities, the χ^2 statistic indicated a poor fit of the model to the data (Pearson's $\chi^2 = 36.3$, df = 14, $P = 0.01$). Further examination of the data indicated that two of the 1765 cells (0.1%) had very large standardized residuals. Both cells were located in reach 9, and although these cells had very low predicted probabilities of containing redds, both were used by spawning fish. Both cells were associated with water depths, water velocities, and stream widths that were highly suitable for spawning but contained substrates that were finer than those typically used by spawning fish (substrate size = 4 and 6 mm) (Fig. 3). Although we had no a priori reason to exclude these cells from the data set (e.g., if the substrate sizes associated with these cells were the result of measurement error), it is nonetheless important to determine whether these two cells alone were responsible for the lack of model fit. When these two "outliers" were removed from the data set, the model fit was greatly improved (Pearson's $\chi^2 = 10.5$, $P = 0.72$).

A visualization of the model fit is provided by the stream channel map of reach 9 shown in Fig. 5. Female golden trout constructed redds in 15 of the 205 cells (7%). The poorer fit

Fig. 3. Shape of the estimated nonparametric function (g) and 95% confidence intervals for each of the highly significant predictors of redd presence or absence: (A) substrate size, (B) water depth, (C) water velocity, and (D) stream width. These functional forms describe the effect of each independent variable on the probability of redd presence. Hatch marks above the x -axes indicate the measured values of each independent variable.

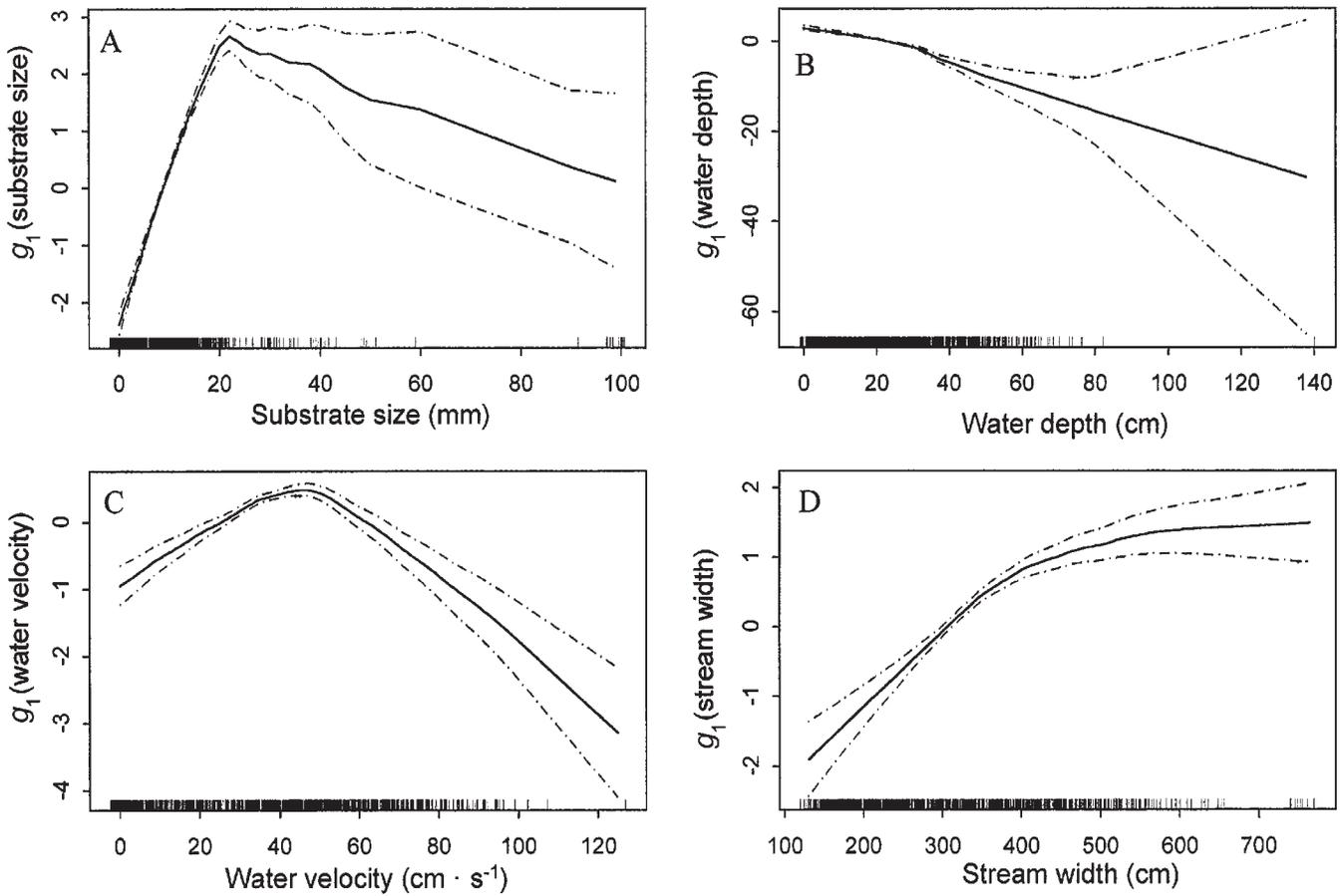
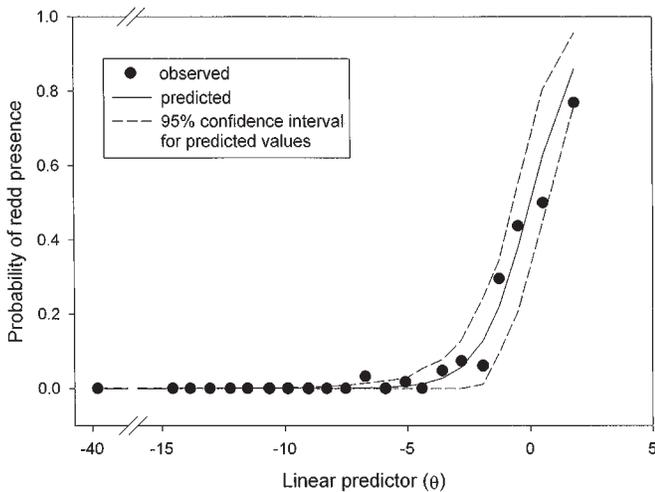


Fig. 4. Results of the cross-validation procedure showing the observed probability of redd presence ($\pm 95\%$ confidence interval) and predicted probability of redd presence for 22 categories of the linear predictor (θ). The close association between the observed and predicted probabilities indicates that the model generally provided a good fit to the data from stream reaches not used in model parameterization.

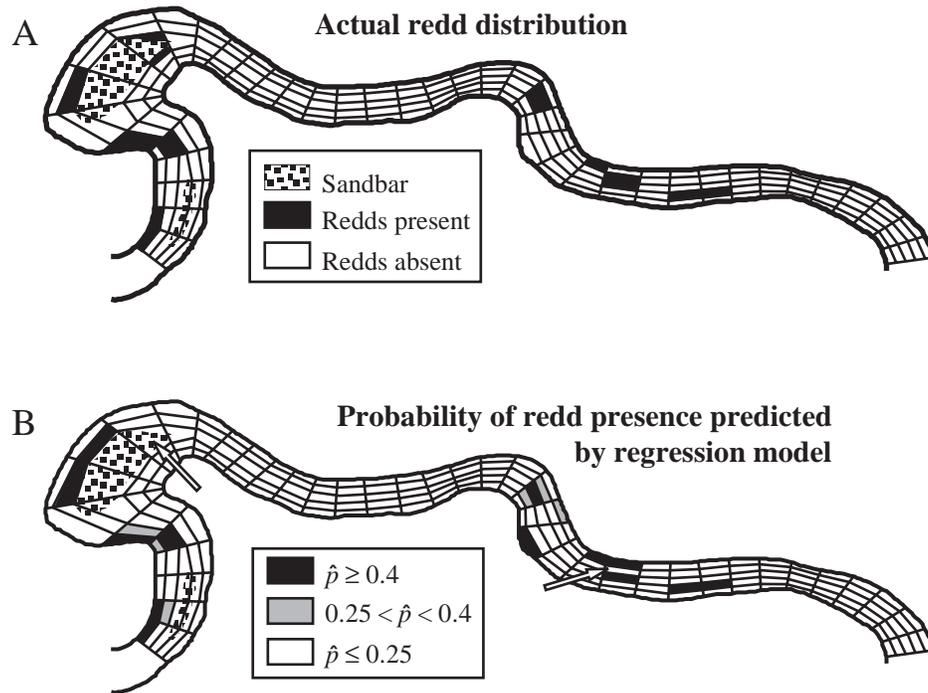


of the model to reach 9 than to all other stream reaches was due to the two outlier cells that contained redds despite having very low predicted probabilities of redd presence ($\hat{p} < 0.2$) (Fig. 5). Despite these outliers, Fig. 5 indicates a close correspondence between the locations of cells with high predicted probabilities of use and the locations of cells that actually contained redds. For cells categorized as $\hat{p} \leq 0.25$, $0.25 < \hat{p} < 0.4$, or $\hat{p} \geq 0.4$, 1.1, 33.3, and 78.6%, respectively, actually contained redds.

Discussion

Our results indicate that locations chosen by spawning golden trout in Mulkey Creek were strongly influenced by water depth, water velocity, substrate size, and stream width. Of these four highly significant independent variables, substrate size, water depth, and water velocity alone explained the majority of the variation in redd location. This finding is in agreement with the results of other studies that also showed the importance of these three variables in discriminating between sites used and not used by spawning salmonids (Rinne 1980; Witzel and MacCrimmon 1983; Parsons and Hubert 1988; Grost et al. 1990; Thurow and King 1994; Knapp and Vredenburg 1996). The significant effect

Fig. 5. Stream channel map of reach 9 showing (A) the locations of redd-containing cells and (B) the predicted probabilities of use by spawning golden trout for each cell, calculated using the parametric logistic regression equation. In Fig. 5B, the locations of two outliers described in the Results section are indicated by arrows. For scale, each cell is about 2.5 m long.



of water velocity in our model is perhaps surprising, since the average water velocity was not significantly different between those cells that were used by spawning golden trout and those that were not. However, the importance of water velocity to spawning trout is suggested by the fact that even though the average water velocity of used and unused cells was not significantly different, water velocities of used cells had a much narrower range than the water velocities of unused cells. This narrower range but similar mean water velocity indicates that spawning females avoided cells with the lowest and highest water velocities.

Stream width explained a relatively small, but highly significant, amount of the variation in redd location. Because the importance of stream width has not been previously reported, some discussion of the role of stream width in influencing spawning locations of golden trout in Mulkey Creek is warranted. Based on the same Mulkey Creek study area, the results of Knapp et al. (1998) indicated that all golden trout redds were constructed in wide stream reaches and they attributed this spatial distribution of redds to the fact that water depths, water velocities, and substrate sizes used by spawning golden trout were much more common in wide than in narrow reaches. While true, the significant effect of stream width in our model even after the effects of water depth, water velocity, and substrate size were accounted for suggests that female golden trout either used stream width (in addition to depth, velocity, and substrate) to select spawning locations or selected spawning locations based in part on additional habitat characteristics that were also associated with wide reaches but were not measured in our study.

Previous studies of spawning habitat preferences by stream salmonids have suggested additional factors beyond

those measured in our study that may influence spawning locations (e.g., groundwater upwelling, river bottom topography; Witzel and MacCrimmon 1983; Shirvell 1989). It could be argued, therefore, that our study failed to provide a complete evaluation of all the important factors influencing choice of spawning location by golden trout. While incorporation of other variables, such as groundwater upwelling, might have increased the explanatory ability of our model, the fact that our models explained nearly 60% of the variation in redd location and accurately predicted the cell-specific probabilities of use indicates that our choice of habitat variables was sufficiently complete. In addition, we purposefully developed our habitat model based exclusively on easily measured variables to facilitate future model application. Because groundwater upwelling and river bottom topography are relatively difficult to quantify, we did not consider these variables in our study.

Our model provided a good fit to all but two of the 1765 cells used in this study, but the two exceptions are worthy of additional discussion. The cross-validation procedure identified two cells located in a single reach that contained redds despite having very low predicted probabilities of redd presence. This discrepancy is likely the result of within-cell variability in habitat characteristics. Although we attempted to minimize within-cell variability by using small cells ($\sim 0.75 \text{ m}^2$), observations made during measurement of habitat characteristics indicated that a small subset of cells still showed considerable within-cell variability. In these cases, habitat characteristics measured in the center of the cell may not have accurately reflected habitat conditions elsewhere in the cell. Because elimination of these two cells from the data set changed a previously poor fit to a very good fit, we conclude that while the parametric model did not adequately fit

data from all 1765 cells, it did provide a very good fit to all 1674 cells that did not contain redds and to 89 of the 91 redd-containing cells (97.8%).

Spatial autocorrelation in data and colinearity between independent variables may be common in ecological studies (Legendre 1993; Thomson et al. 1996), including those investigating fish–habitat relationships (Mathur et al. 1985), and can pose substantial problems for regression analyses. In our study, spatial autocorrelation could result in an underestimation of the uncertainty associated with parameter estimates by reducing the independence of neighboring cells. We surmounted this problem by including the nonlinear spatial variable “cell location,” following the methods of Legendre (1993) and Hobert et al. (1997). The fact that cell location was only marginally significant and increased the explained variation by only 1.6% suggests either that spatial dependence was relatively unimportant in understanding the factors influencing the choice of spawning location by female golden trout or that the habitat variables included in the model accounted for most of the spatial autocorrelation between neighboring cells. Colinearity between independent variables will cause the regression coefficient of any independent variable to depend on which other independent variables are included in the model (Neter et al. 1990). Because of the likely presence of correlations between independent variables in our model, the regression coefficients might change if the independent variables in the model were changed (e.g., by the addition of a new independent variable). Therefore, we interpreted the regression coefficients only in the presence of all other variables in the model.

Given the widely held belief that locations chosen by spawning stream salmonids are strongly influenced by water depth, water velocity, and substrate size (Bjornn and Reiser 1991), it is surprising that the single study that we were able to find that attempted to predict spawning locations based on these habitat characteristics indicated a very poor match between locations predicted as suitable for spawning and locations actually used by spawning fish (Shirvell 1989). This failure is in marked contrast with the ability of our model to accurately predict the cell-specific probability of use by spawning golden trout.

Our use of a GAM is a substantial improvement over previous approaches to modeling fish–habitat relationships. Past studies have relied on generalized linear models such as traditional logistic regression and have therefore needed to (i) assume a linear relationship between the logit of the dependent variable and each of the habitat variables (Thielke 1985), (ii) transform continuous variables into categorical variables (Thielke 1985), or (iii) incorporate nonlinear terms based on an iterative process that is blind to the actual shapes of the response curves (Bozek and Rahel 1991). Our use of a GAM allowed us to assess the significance of habitat variables without making assumptions about the shapes of the relationships between the independent variables and the probability of use. After using plots of the estimated nonparametric functions to evaluate the form of each significant relationship, we constructed a parametric logistic regression model that incorporated nonlinear terms chosen explicitly to approximate each relationship.

The results of our study indicate that the probabilities of golden trout in Mulkey Creek using a particular location for

spawning are predictable based on easily measured microhabitat and channel characteristics. Since the habitats used by spawning golden trout in Mulkey Creek are qualitatively similar to habitats used by spawning golden trout in other streams (R.A. Knapp, personal observation) and to habitats used by several other salmonid species (Knapp and Vredenburg 1996), it seems likely that habitat use by spawning salmonids will generally be predictable using a model approach similar to the one that we used in this study. Further research will be necessary, however, to determine how generally applicable our particular model is to golden trout populations outside our study area and to other species of stream-dwelling salmonids.

In conclusion, although nonparametric regression approaches have only recently begun to be used in ecological studies (e.g., Wiser et al. 1998), our results suggest that they may have considerable utility in the development of models to predict fish distributions based on habitat characteristics. Given the escalating pace at which fish habitats are being altered, such models are increasingly important in predicting the effects of these alterations on populations.

Acknowledgements

R.A. Knapp was supported during this research by Cooperative Agreement PSW-95-0018CA between the U.S. Forest Service Pacific Southwest Research Station and the Marine Science Institute, University of California, Santa Barbara. R.A. Knapp thanks K. Matthews for financial and logistic support during this research and V. Vredenburg and S. Miller for their able assistance with the field work. We also thank two anonymous reviewers whose comments on an earlier version of the manuscript greatly improved this paper.

References

- Beecher, H.A., Johnson, T.H., and Carleton, J.P. 1993. Predicting microdistributions of steelhead (*Oncorhynchus mykiss*) parr from depth and velocity preference criteria — test of an assumption of the Instream Flow Incremental Methodology. *Can. J. Fish. Aquat. Sci.* **50**: 2380–2387.
- Bjornn, T.C., and Reiser, D.W. 1991. Habitat requirements of salmonids in streams. *Am. Fish. Soc. Spec. Publ.* **19**: 83–138.
- Bovee, K.D. 1978. The incremental method of assessing habitat potential for coolwater species, with management implications. *Am. Fish. Soc. Spec. Publ.* **11**: 340–346.
- Bozek, M.A., and Rahel, F.J. 1991. Assessing habitat requirements of young Colorado River cutthroat trout by use of macrohabitat and microhabitat analyses. *Trans. Am. Fish. Soc.* **120**: 571–581.
- Breiman, L., Friedman, J.H., Olshen, R., and Stone, C.J. 1984. *Classification and regression trees*. Wadsworth, Belmont, Calif.
- California Department of Fish and Game. 1996. Instream flow and habitat development investigations for Mill Creek, Mono County, California. Stream Evaluation Rep. 96-1, California Department of Fish and Game, Sacramento, Calif.
- Cleveland, W.S., and Devlin, S.J. 1988. Locally weighted regression: an approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **83**: 596–610.
- Efron, B., and Tibshirani, R. 1991. *Statistical analysis in the computer age*. Science (Washington, D.C.), **253**: 390–395.

- Grost, R.T., Hubert, W.A., and Wesche, T.A. 1990. Redd site selection by brown trout in Douglas Creek, Wyoming. *J. Freshwater Ecol.* **5**: 365–371.
- Hobert, J.P., Altman, N.S., and Schofield, C.L. 1997. Analyses of fish species richness with a spatial covariate. *J. Am. Stat. Assoc.* **92**: 846–854.
- Hosmer, D.W., Jr., and Lemeshow, S. 1989. *Applied logistic regression*. John Wiley & Sons, New York.
- Knapp, R.A., and Vredenburg, V.T. 1996. Spawning by California golden trout: characteristics of spawning fish, seasonal and daily timing, redd characteristics, and microhabitat preferences. *Trans. Am. Fish. Soc.* **125**: 519–531.
- Knapp, R.A., Vredenburg, V.T., and Matthews, K.R. 1998. Effects of stream channel morphology on golden trout spawning habitat and recruitment. *Ecol. Appl.* **8**: 1104–1117.
- Legendre, P. 1993. Spatial autocorrelation: trouble or a new paradigm? *Ecology*, **74**: 1659–1673.
- Linhart, H., and Zucchini, W. 1986. *Model selection*. John Wiley & Sons, New York.
- Lorenz, J.M., and Eiler, J.H. 1989. Spawning habitat and redd characteristics of sockeye salmon in the glacial Taku River, British Columbia and Alaska. *Trans. Am. Fish. Soc.* **118**: 495–502.
- MathSoft. 1997. *S-Plus user's manual, version 4*. Data Analysis Products Division, MathSoft, Seattle, Wash.
- Mathur, D., Bason, W.H., Purdy, E.J., Jr., and Silver, C.A. 1985. A critique of the Instream Flow Incremental Methodology. *Can. J. Fish. Aquat. Sci.* **42**: 825–831.
- McCullagh, P., and Nelder, J.A. 1989. *Generalized linear models*. Chapman and Hall, London, U.K.
- Nagelkerke, N.J.D. 1991. A note on a general definition of the coefficient of determination. *Biometrika*, **78**: 691–692.
- Nehring, R.B., and Anderson, R.M. 1993. Determination of population-limiting critical salmonid habitats in Colorado streams using the Physical Habitat Simulation System. *Rivers*, **4**: 1–19.
- Neter, J., Wasserman, W., and Kutner, M.H. 1990. *Applied linear statistical models: regression, analysis of variance, and experimental designs*. Irwin, Homewood, Ill.
- Parsons, B.G.M., and Hubert, W.A. 1988. Influence of habitat availability on spawning site selection by kokanees in streams. *N. Am. J. Fish. Manage.* **8**: 426–431.
- Platts, W.S., Megahan, W.F., and Minshall, G.W. 1983. *Methods for evaluating stream, riparian, and biotic conditions*. U.S. For. Serv. Gen. Tech. Rep. INT-138.
- Reiser, D.W., Wesche, T.A., and Estes, C. 1989. Status of instream flow legislation and practices in North America. *Fisheries (Bethesda)*, **14**: 22–29.
- Rinne, J.N. 1980. Spawning habitat and behavior of Gila trout, a rare salmonid of the southwestern United States. *Trans. Am. Fish. Soc.* **109**: 83–91.
- Rosgen, D.L. 1994. A classification of natural rivers. *Catena*, **22**: 169–199.
- Shirvell, C.S. 1989. Ability of PHABSIM to predict chinook salmon spawning habitat. *Regul. Rivers Res. Manage.* **3**: 277–289.
- Stalnaker, C.B. 1979. The use of habitat structure preferences for establishing flow regimes necessary for maintenance of fish habitat. *In The ecology of regulated streams. Edited by J.V. Ward and J.A. Stanford*. Plenum Press, New York. pp. 321–337.
- Stefferd, J.A. 1993. Spawning season and microhabitat use by California golden trout (*Oncorhynchus mykiss aguabonita*) in the southern Sierra Nevada. *Calif. Fish Game*, **79**: 133–144.
- Thielke, J. 1985. A logistic regression approach for developing suitability-of-use functions for fish habitat. *In Proceedings of the Symposium on Small Hydropower and Fisheries. Edited by F.W. Olson, R.G. White, and R.H. Hamre*. American Fisheries Society, Aurora, Co. pp. 32–38.
- Thomson, J.D., Weiblen, G., Thomson, B.A., Alfaro, S., and Legendre, P. 1996. Untangling multiple factors in spatial distributions: lilies, gophers, and rocks. *Ecology*, **77**: 1698–1715.
- Thurrow, R.F., and King, J.G. 1994. Attributes of Yellowstone cutthroat trout redds in a tributary of the Snake River, Idaho. *Trans. Am. Fish. Soc.* **123**: 37–50.
- Venables, W.N., and Ripley, B.D. 1994. *Modern applied statistics with S-Plus*. Springer-Verlag, New York.
- Wiser, S.K., Peet, R.K., and White, P.S. 1998. Prediction of rare-plant occurrence: a Southern Appalachian example. *Ecol. Appl.* **8**: 909–920.
- Witzel, L.D., and MacCrimmon, H.R. 1983. Redd-site selection by brook trout and brown trout in southwestern Ontario streams. *Trans. Am. Fish. Soc.* **112**: 760–771.
- Yee, T.W., and Mitchell, N.D. 1991. Generalized additive models in plant ecology. *J. Veg. Sci.* **2**: 587–602.