North American Journal of Fisheries Management

ISSN: 0275-5947 (Print) 1548-8675 (Online) Journal homepage: http://www.tandfonline.com/loi/ujfm20

Predictive Models of Fish Species Distribution in the Blackwater Drainage, British Columbia

Marc S. Porter , Jordan Rosenfeld & Eric A. Parkinson

To cite this article: Marc S. Porter , Jordan Rosenfeld & Eric A. Parkinson (2000) Predictive Models of Fish Species Distribution in the Blackwater Drainage, British Columbia, North American Journal of Fisheries Management, 20:2, 349-359, DOI: 10.1577/1548-8675(2000)020<0349:PMOFSD>2.3.CO;2

To link to this article: http://dx.doi.org/10.1577/1548-8675(2000)020<0349:PMOFSD>2.3.CO;2

Published online: 08 Jan 2011.

North American Journal of Fisheries Management 20:349–359, 2000

Predictive Models of Fish Species Distribution in the Blackwater Drainage, British Columbia

MARC S. PORTER,* JORDAN ROSENFELD, AND ERIC A. PARKINSON

British Columbia Fisheries Branch, Research and Development Section, University of British Columbia, 2204 Main Mall, Vancouver, British Columbia V6T 1Z4, Canada

Abstract.—Management of fish biodiversity and individual fish species requires the ability to understand and predict expected species distributions. Models predicting species distributions can provide insight into habitat requirements, help identify unique outlier populations, and determine potential biodiversity hotspots. Our goal was to determine whether reliable models of species distributions could be developed for freshwater fish in British Columbia using large-scale macrohabitat data. We surveyed 48 stream sites in a British Columbia drainage with high species diversity (the Blackwater River) and developed statistical models based on macrohabitat variables to describe fish species distributions. Classification rates (i.e., the proportion of sites correctly classified by the model) of our logistic regression models based solely on map-based variables were generally high (73–90%) for most fish species found in the Blackwater River drainage. Including field-measured variables produced significantly better models for most species, but improvements in classification rates were generally marginal. Application of the Blackwater River species models to data from a geographically distant drainage (the Similkameen River) was not successful (i.e., classification rates were poor for all shared species). Classification rates for species in the Similkameen River drainage were improved considerably by using the same macrohabitat variables to generate logistic models that were unique to that watershed. Combining data for the two watersheds also generated significant logistic regressions. Quality of these combined models was significantly improved for most species by incorporating a categorical variable that presumably captured broad differences in habitat conditions between the watersheds; however, classifications were poorer than models specific to individual drainages. Further refinements to quantify variation in habitat conditions among watersheds should permit the development of regional fish distribution models as a layer in geographic information systems. In conjunction with map-based macrohabitat information, distribution models could provide a powerful diagnostic and predictive tool for improving watershed planning and management practices.

Conserving fish biodiversity requires reliable information on distribution and habitat use by individual fish species. This is especially true for species that are at risk because they exist in low numbers, are endemic or disjunctively distributed, or are sensitive to the effects of agriculture, urbanization, and forestry (McPhail and Carveth 1992; Cannings and Ptolemy 1998). Although distributions and habitat requirements are known for most game species in British Columbia, those of many nongame species are poorly documented. Therefore, systematic definitions of fish species distributions and habitat use are critical if management of fish and fish habitat are to be effectively integrated into land-use planning (Parkinson et al. 1996).

The identification of habitat features that characterize fish distributions and abundance has long been an objective of fisheries research, but the focus has largely been on salmonids (e.g., Bjornn and Reiser 1991; Meehan and Bjornn 1991). Much of this work has attempted to characterize habitats through small-scale approaches such as stream reach or habitat unit inventories (e.g., Fausch and Northcote 1992; Nickelson et al. 1992). It may be more important, for planning purposes, to determine fish distributions and responses to habitat based on a larger-scale geomorphic approach (Nelson et al. 1992; Rieman and McIntyre 1995; Kruse et al. 1997). Previous research has shown that large-scale geomorphic variables (e.g., elevation, channel slope, stream size) can be used as reliable predictors of fish species occurrences (Gorman and Karr 1978; Lanka et al. 1987; Nelson et al. 1992; Rieman and McIntyre 1995; Kruse et al. 1997; Watson and Hillman 1997). The development of incidence functions based on large-scale variables could be an important tool for fisheries and forestry managers, allowing them to predict the presence or absence of particular fish species, even where little detailed habitat or population information is available (Rieman and McIntyre 1995; Kruse et al. 1997). Models predicting species dis-

* Corresponding author: mporter@ubc.env.gov.bc.ca

Received September 8, 1998; accepted November 11, 1999

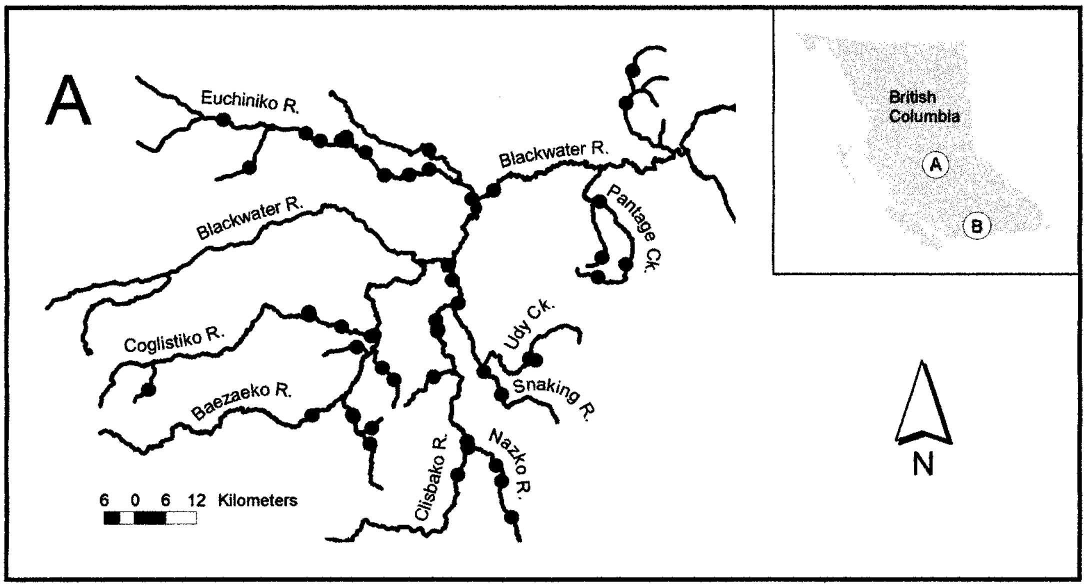

FIGURE 1.—Location of stream sites surveyed within the Blackwater River drainage. Inset indicates the locations of the Blackwater River drainage (A) and the Similkameen River drainage (B) in British Columbia.

tribution can also identify habitat requirements, unique outlier populations, and potential biodiversity hotspots. Although predictions of presence–absence will need to be verified with ground surveys, predictive models can help focus inventory and management activities on areas where species are considered likely to occur and may allow managers to anticipate the possible consequences of large-scale habitat change.

Our purpose was to quantify macrohabitat use by individual fish species in the Blackwater River drainage, a representative high-diversity watershed in central British Columbia, and to identify the physical variables that are most important in discriminating between stream sites with and without individual fish species. Our primary objective was to predict the presence of different fish species at a site and to develop predictive habitat models that are suitable for interfacing with GIS-based planning and information systems. We developed logistic regression models using a combination of map-based features (e.g., watershed gradient, drainage area) and data collected in the field (e.g., bankfull width, stream temperature) to determine the level of information necessary to reliably predict fish species distributions. We then tested the derived relationships with data collected independently from a distant watershed (the Similkameen River drainage), to determine the utility of such models for management purposes.

Study Area

The Blackwater River is on the Fraser Plateau of British Columbia's central interior (Figure 1) and drains an area of approximately 12,400 km^2 (Water Survey of Canada 1989). Climate in the area is dry for British Columbia; winters are cold and summers short (Griffith 1991). The existing forest in the drainage is predominated by lodgepole pine *Pinus contorta*, white spruce *Picea glauca*, and black spruce *Picea mariana*; Interior douglas fir *Pseudotsuga menziesii* is included in the wetter areas. Varied degrees of timber harvesting (including extensive clear-cut logging) and limited agriculture in valley bottoms occur throughout much of the Blackwater drainage. At least 19 species of fish are known to occur within the Blackwater drainage, making it an ideal watershed to assess habitat use by multiple species.

Methods

Data collection.—Fish were sampled during summer low-flow periods (July–September) at 36 sites in 1996 and 12 sites in 1997 (Figure 1). We surveyed a mixture of small, intermediate, and large streams to cover the full range of flowing-water habitats in the drainage. Sampling sites were of variable size but included all naturally occurring habitat units (e.g., riffles, pools, runs, glides, cascades, etc.) present within a stream reach (defined

TABLE 1.—Physical habitat variables used to characterize sites in the Blackwater River drainage, British Columbia. Map-based variables were estimated from 1:50,000 topographic maps. Field-based variables were estimated from measurements at a site by a field crew in 1996 and 1997.

Variable	Type	Description
Drainage area (km^2)	Map-based	Total area (± 1 km^2) from which water drains into the stream upstream of the survey site
Watershed gradient (%)	Map-based	Drop in elevation over length of stream from headwaters to site of sampling ($\pm 0.5\%$)
Reach gradient (%)	Map-based	Drop in elevation ($\pm 0.5\%$) over length of reach encompassing survey site (defined by 30.5-m (100-ft) contour change bracketing site)
Elevation (m)	Map-based	Site elevation (± 15 m) as determined from topographic maps or handheld Geographic Positioning Systems units or both
Distance from a lake >4 ha (km)	Map-based	Linear distance (± 0.1 km) from site to nearest large lake, following the natural meanders of the stream
Distance from the Fraser River (km)	Map-based	Linear distance (± 0.1 km) from site to the nearest major river (Fraser River), following the natural meanders of the stream
Lake area in drainage (ha)	Map-based	Total surface area (± 0.1 ha) of all lakes present within the defined drainage area
Sinuosity	Map-based	Channel meander length over down-valley linear distance
Site gradient (%)	Field-based	Drop in water-surface elevation ($\pm 0.5\%$) over total length of stream channel surveyed
Bank-full width (m)	Field-based	Average width (± 0.1 m) of the bank-full channel (defined by the presence of permanent rooted vegetation)
Bank-full depth (cm)	Field-based	Average depth (± 1 cm) from the height of the bank-full channel to the channel bottom
Maximum wetted depth (cm)	Field-based	Deepest point (± 1 cm) in stream channel
Maximum temperature (°C)	Field-based	Highest temperature (± 0.1°C) recorded at site by hourly temperature-logger readings
Degree-days (°C·days)	Field-based	Area under curve of mean daily temperature versus time over a single year

in Johnston and Slaney 1996). Fish were collected at all sites using a Smith–Root backpack electroshocker. At deeper sites 10-m-long sinking gill nets with 7.5-cm, 5.0-cm, 2.5-cm, and multipanel (1.25–3.25 cm) mesh sizes were set parallel to the current. Gill nets were generally fished for 2–4 h and were checked regularly during the set to reduce fish mortality. Minnow traps, seines (both pole and beach), trap nets, and angling were also used where appropriate. All or some of these sampling methods were employed to ensure a high rate of capture success for all species present at a particular site. Captured fish were identified to species in the field, measured in fork length (mm), weighed to the nearest 0.1 g, and released. Unidentifiable fish were measured, preserved in 10% formalin, and later identified in the laboratory. Species not captured after extensive sampling with the full range of sampling techniques were assumed to be absent from a site.

Physical characteristics of each site were measured based on procedures outlined in Newbury and Gaboury (1994) and Johnston and Slaney (1996). Fourteen habitat variables (Table 1) were estimated from 1:50,000 topographic maps or by direct measurement in the field. Site habitat variables (e.g., bankfull width, drainage area, gradient measures) were those previously identified as useful indicators of fish species occurrence (Lanka et al. 1987; Rieman and McIntyre 1995; Kruse et al. 1997; Watson and Hillman 1997), many of which are also correlated with habitat features at a finer scale (e.g., current velocity, substrate type, channel stability, pool–riffle frequencies) that are known to influence fish distribution and abundance (Gorman and Karr 1978; Angermeier and Karr 1983; Moyle and Baltz 1985; Lobb and Orth 1991; Hubert and Kozel 1993).

Data analysis.—Statistical models that could predict presence–absence of individual species were developed using binary logistic regression. Logistic regression was used because of the binomial nature of the dependent variable (presence–absence) and because the analysis provides a probabilistic prediction of fish presence. The analysis is insensitive to differences in fish densities at sites and was biased only if a fish species present at a stream site was never detected. We believe we minimized this possibility by sampling each site thoroughly and by using a suite of different sampling techniques for both juvenile and adult fish. Logistic regression analysis, unlike most other mul-

TABLE 2.—Means (± SD) of principal habitat measurements for sites in British Columbia's Blackwater River drainage at which individual fish species were considered to be present or absent based on summer sampling in 1996 and 1997.

Species	Status	N[a]	Drainage area (km^2)	Watershed gradient (%)	Reach gradient (%)	Elevation (m)
Bridgelip sucker	Present	18	2,889 ± 3,419	1.0 ± 0.5	0.7 ± 0.8	870 ± 64
Catostomus commersoni	Absent	30	593 ± 1,320	2.0 ± 1.3	1.3 ± 1.0	937 ± 111
Burbot	Present	19	2,306 ± 2,852	0.9 ± 0.5	0.8 ± 0.8	876 ± 73
Lota lota	Absent	29	895 ± 2,230	2.1 ± 1.3	1.3 ± 1.1	935 ± 110
Chinook salmon	Present	19	2,492 ± 3,515	1.2 ± 0.6	1.2 ± 1.1	885 ± 85
Oncorhynchus tshawytscha	Absent	29	773 ± 1,366	1.9 ± 1.4	1.0 ± 0.9	930 ± 107
Chiselmouth	Present	14	3,635 ± 3,561	0.9 ± 0.5	0.6 ± 0.4	836 ± 33
Arcocheilus alutaceus	Absent	34	556 ± 1,227	1.9 ± 1.2	1.3 ± 1.1	943 ± 103
Largescale sucker	Present	14	3,292 ± 3,726	0.9 ± 0.5	0.5 ± 0.4	839 ± 37
Catostomus macrocheilus	Absent	34	697 ± 1,363	1.9 ± 1.3	1.3 ± 1.1	942 ± 103
Leopard dace	Present	14	3,610 ± 3,579	1.0 ± 0.5	0.5 ± 0.4	837 ± 31
Rhinichthys falcatus	Absent	34	566 ± 1,233	1.9 ± 1.3	1.3 ± 1.1	943 ± 103
Longnose dace	Present	19	2,260 ± 2,862	1.0 ± 0.6	1.0 ± 1.1	888 ± 75
R. cataractae	Absent	29	925 ± 2,242	2.0 ± 1.3	1.2 ± 1.0	928 ± 113
Longnose sucker	Present	17	1,452 ± 3,092	1.6 ± 1.3	1.0 ± 0.9	896 ± 91
Catostomus catostomus	Absent	31	1,455 ± 2,278	1.6 ± 1.2	1.1 ± 1.1	921 ± 106
Mountain whitefish	Present	26	2,340 ± 3,141	1.0 ± 0.5	1.0 ± 1.0	892 ± 86
Prosopium williamsoni	Absent	22	406 ± 909	2.3 ± 1.4	1.2 ± 1.0	936 ± 112
Northern squawfish	Present	19	2,861 ± 3,357	0.9 ± 0.5	0.5 ± 0.4	833 ± 38
Ptychocheilus oregonensis	Absent	29	532 ± 1,246	2.0 ± 1.3	1.5 ± 1.1	964 ± 95
Peamouth	Present	13	3,227 ± 3,728	1.0 ± 0.6	0.6 ± 0.4	837 ± 30
Mylocheilus caurinus	Absent	35	795 ± 1,588	1.8 ± 1.3	1.3 ± 1.1	940 ± 104
Prickly sculpin	Present	16	1,419 ± 1,299	0.9 ± 0.5	0.5 ± 0.4	854 ± 42
Cottus asper	Absent	32	1,471 ± 3,023	2.0 ± 1.3	1.4 ± 1.1	941 ± 109
Rainbow trout	Present	40	1,643 ± 2,718	1.3 ± 0.8	1.1 ± 1.0	911 ± 102
O. mykiss	Absent	8	349 ± 726	3.2 ± 1.9	1.3 ± 1.0	919 ± 97
Redside shiner	Present	22	2,664 ± 3,172	1.1 ± 1.0	0.5 ± 0.4	853 ± 57
Richardsonius balteatus	Absent	26	430 ± 1,231	2.0 ± 1.2	1.6 ± 1.1	962 ± 103

[a] Number of sites species was present or absent.

tivariate procedures, does not require that variables be normally distributed, linearly related, or of equal variance within each group (Press and Wilson 1978; Tabachnick and Fidell 1996). With two possible outcomes (in this case, presence or absence) the form of the logistic function is

$$P = e^u/(1 + e^u),$$

where P = the estimated probability, e = the inverse natural logarithm of 1, and u = linear model; this model is

$$u = A + b_1X_1 + b_2X_2 + \ldots + b_nX_m,$$

where A = regression constant, b_n = regression coefficients, and X_m = independent variables.

Before generating logistic regression models, Pearson correlation matrices were calculated to test for multicollinearity among the site variables. Multicollinearity occurs when variables are too highly correlated, which can result in both logical and statistical problems in a multivariate analysis (Tabachnick and Fidell 1996). Because a number of the variables were highly correlated with each other, we selected a subset of variables for use in logistic regression analysis that (1) were relatively uncorrelated ($|r| < 0.7$), and (2) had the highest correlation with excluded variables (as in Eadie and Keast 1982). Four map-based variables (drainage area, watershed gradient, reach gradient, and elevation) and two field-based variables (bankfull width and maximum temperature) were identified as independent habitat dimensions using this screening process.

To illustrate univariate differences between sites with and without individual fish species, mean values (±SD) were calculated for each of the selected habitat variables. Logistic regression models were fit for both untransformed and log-transformed variables; both sets of variables produced similar results, but the untransformed data produced better-fitting models and are the only results presented here. We used SYSTAT LOGIT (Steinberg and Colla 1991) for all logistic regression analyses.

Statistical significance of the models was assessed using a chi-square test ($\alpha = 0.05$). The

TABLE 2.—Extended.

Species	Bank-full width (m)	Maximum temperature (°C)
Bridgelip sucker	31.0 ± 15.4	20.8 ± 1.5
Catostomus commersoni	11.9 ± 19.3	19.4 ± 1.8
Burbot	25.5 ± 15.1	20.6 ± 1.6
Lota lota	14.9 ± 21.9	19.7 ± 1.9
Chinook salmon	20.2 ± 16.9	19.6 ± 1.6
Oncorhynchus tshawytscha	18.3 ± 22.1	20.6 ± 1.9
Chiselmouth	39.5 ± 22.7	21.2 ± 1.0
Arcocheilus alutaceus	10.6 ± 10.9	19.4 ± 1.9
Largescale sucker	39.6 ± 22.7	21.4 ± 0.9
Catostomus macrocheilus	10.6 ± 10.8	19.3 ± 1.7
Leopard dace	38.3 ± 22.9	21.2 ± 1.0
Rhinichthys falcatus	11.2 ± 12.0	19.4 ± 1.8
Longnose dace	24.2 ± 15.0	20.5 ± 1.5
R. cataractae	15.7 ± 22.4	19.7 ± 2.1
Longnose sucker	17.3 ± 17.8	20.6 ± 1.4
Catostomus catostomus	20.0 ± 21.4	19.8 ± 2.0
Mountain whitefish	28.5 ± 21.6	20.5 ± 1.6
Prosopium williamsoni	7.9 ± 10.1	19.2 ± 1.9
Northern squawfish	34.0 ± 22.4	21.4 ± 0.9
Ptychocheilus oregonensis	9.3 ± 10.0	18.9 ± 1.6
Peamouth	33.9 ± 16.9	21.0 ± 1.4
Mylocheilus caurinus	13.5 ± 18.4	19.6 ± 1.8
Prickly sculpin	33.9 ± 23.8	21.3 ± 1.4
Cottus asper	11.6 ± 12.8	19.2 ± 1.6
Rainbow trout	20.4 ± 20.5	20.1 ± 1.8
O. mykiss	11.2 ± 15.8	19.8 ± 0.5
Redside shiner	32.3 ± 22.3	21.0 ± 1.3
Richardsonius balteatus	7.9 ± 7.3	18.9 ± 1.7

goodness of fit of each model was evaluated by the magnitude of McFadden's rho-square, ρ^2 (Steinberg and Colla 1991), a summary statistic indicating the accuracy with which the model fits the observed data (analogous to the R^2 in a linear regression model). Rho-square values are always between 0 and 1, and a higher ρ^2 corresponds to more significant results. Due to the binary nature of the dependent variable in logistic regressions, however, correlations between predictors and the dependent variable are much lower than would be obtained in a linear regression. Consequently, ρ^2-values obtained may seem quite low, yet still represent a good model fit. Rho-square values between 0.20 and 0.40, for example, are considered very satisfactory (Steinberg and Colla 1991). Models were also assessed using the percent classification (the proportion of sites correctly classified by the model), a more readily interpretable criteria. Probabilities for species occurrence were calculated from the logistic models and compared with observed values. Predicted probabilities of 0.50 or greater indicated species presence whereas probabilities less than 0.50 indicated species absence. A high correspondence between predicted and observed presence–absence indicated a good model fit. Classification rates assigned in this manner are somewhat biased because the classified sites were also used to estimate model parameters; consequently, true error rates for the species models are probably higher.

Two sets of logistic regression models were constructed for each species, one using solely map-based variables (drainage area, watershed gradient, reach gradient, and elevation) and a second incorporating the map-based variables and habitat variables collected in the field (maximum stream temperature and bankfull width). Our objective was to assess the relative predictive ability of models based on map-based versus combined map-and-field-based variables. A number of different map-and-field-based models were produced using combinations of the different field-collected variables. Improvement in model quality was formally assessed by a likelihood ratio test (*G* statistic) that computes the difference in log-likelihoods for any pair of nested models (Steinberg and Colla 1991).

To test model performance in predicting fish species occurrences, logistic regressions generated from our surveys were applied to an independent data set derived from fish and habitat surveys in a geographically distant watershed (the Similkameen River; Rosenfeld 1996). The Similkameen drainage is another high-diversity watershed in southwestern British Columbia (Figure 1), sharing at least seven fish species with the Blackwater drainage. Sampling methods used at sites in the Similkameen drainage were similar to those employed in the Blackwater drainage and were considered sufficient to be reflective of fish species presence or absence. Habitat information from these surveys was used to test map-based (habitat information derived from 1:50,000 topographic maps) and map-and-field-based logistic models. Predicted probabilities of 0.5 or greater indicated probable fish presence, whereas probabilities less than 0.5 indicated absence. The reliability of individual habitat predictors across regions was additionally tested by generating logistic regression models for the Similkameen data set by using the same habitat variables employed in the Blackwater models, with one exception. Maximum stream temperatures determined from year-long sets of data-loggers in the Blackwater drainage were not available in the Similkameen drainage where a single measurement of stream temperature was taken at time of summer sampling. The quality of model

TABLE 3.—Summary of logistic regression models describing fish species presence and absence in the Blackwater River drainage, British Columbia. Models are based either solely on map-derived variables or a combination of map- and field-derived variables estimated in 1996 and 1997. The significance (*P*-value) of the model, McFadden's rho-squared (ρ^2), model classification rate (percent present, absent, overall), and percentages of sites where individual species were present and absent (in parentheses) are given for each species. Models incorporating field-derived variables are presented only if the likelihood ratio test (*G*-statistic) indicated a significant improvement over the species' map-based model.

				Classification rate (%) and [% of sites]		
Species	Model type	*P*	McFadden's ρ^2	Present	Absent	Overall
Bridgelip sucker	Map-based	0.002	0.276	72 (38)	87 (63)	81
Burbot	Map-based	0.003	0.246	68 (40)	76 (60)	73
Chinook salmon	Map-based	0.013	0.196	53 (40)	86 (60)	73
	Map- and field-based[a]	0.023	0.295	65	74	69
Chiselmouth	Map-based	0.000	0.475	79 (29)	88 (71)	85
	Map- and field-based[b]	0.000	0.603	79	88	85
Largescale sucker	Map-based	0.000	0.399	64 (29)	85 (71)	79
	Map- and field-based[a]	0.000	0.554	86	77	81
Leopard dace	Map-based	0.000	0.486	79 (29)	94 (71)	90
	Map- and field-based[b]	0.000	0.559	71	91	85
Longnose dace	Map-based	0.006	0.224	63 (40)	83 (60)	75
Longnose sucker	Map-based	0.877	N/A	N/A	N/A	N/A
Mountain whitefish	Map-based	0.000	0.417	92 (54)	82 (46)	88
	Map- and field-based[b]	0.000	0.545	96	86	92
Northern squawfish	Map-based	0.000	0.677	89 (40)	90 (60)	90
	Map- and field-based[b]	0.000	0.750	95	93	94
Peamouth	Map-based	0.001	0.332	62 (27)	89 (73)	81
	Map- and field-based[a]	0.005	0.393	85	83	83
Prickly sculpin	Map-based	0.000	0.410	75 (33)	81 (67)	79
	Map- and field-based[c]	0.000	0.835	94	95	94
Rainbow trout	Map-based	0.002	0.425	98 (83)	43 (17)	90
Redside shiner	Map-based	0.000	0.468	82 (46)	81 (54)	81
	Map- and field-based[a]	0.001	0.553	86	89	88

[a] Best combined model based on inclusion of bankfull width and maximum stream temperature.
[b] Best combined model based on inclusion of bankfull width only.
[c] Best combined model based on inclusion of maximum stream temperature only.

fit for these new regressions was assessed by classification rates for each sampled fish species.

Finally, combined logistic regression models were constructed for fish species based on map-based habitat information from both the Similkameen and Blackwater drainages. The objective was to assess whether a common regression for each species, based on remotely derived information, could adequately describe presence and absence across both drainages. We then incorporated a categorical predictor for ''drainage'' (Blackwater or Similkameen) in the combined model to determine if model fit could be improved by including a generalized artificial variable that might capture some of the inherent variation between watersheds.

Results

Habitat models were constructed for 14 species from a total of 5,190 fish collected at 48 stream sites in the Blackwater drainage in 1996 and 1997 (Table 2). Three additional species (lake chub *Couesius plumbeus,* Pacific lamprey *Lampetra tridentata,* and bull trout *Salvelinus confluentus*) were collected from too few sites ($N < 5$) to perform meaningful analyses. At least one of the measured habitat variables differed considerably between sites with and without fish for all species (Table 2), except for longnose sucker. Significant logistic regression models based on the four independent map-based habitat variables (drainage area, watershed gradient, reach gradient, and elevation) were produced for each of the fish species surveyed in the Blackwater drainage, again with the exception of longnose sucker (Table 3). The overall classification rates of map-based models ranged from a minimum of 73% for burbot and chinook salmon to 90% for leopard dace, northern squawfish, and rainbow trout. Overall classification rates averaged 82% ($\pm$6% SD) for the 13 fish species that generated significant regressions, and the average ρ^2-value for these models was 0.387 ($\pm$0.132 SD).

Of 13 map-based models, 11 were better at classifying sites without an individual species than sites with the species. This difference in classifying sites with species absent versus present

TABLE 4.—Overall correct classification rates of fish species occurrence (presence or absence) at sites surveyed in British Columbia's Similkameen River drainage in 1995. Results are presented for logistic regression models developed for the Blackwater River drainage and for models developed using only fish and habitat information from the Similkameen River drainage (but using habitat predictors analogous to those in the Blackwater regressions).

		Overall classification rates of Similkameen sites (%)						
Model	*N*	Prickly sculpin	Largescale sucker	Longnose dace	Mountain whitefish	Northern squawfish	Rainbow trout	Redside shiner
Blackwater River								
Map-based	46	65	72	67	48	72	65	72
Map- and field-based	46	70	73		53	73		70
Similkameen River								
Map-based	46	93	91	80	83	93	78	85
Map- and field-based	46	98	98		85	93		85

ranged from as little as 1% for northern squawfish to as much as 33% for chinook salmon. Models for only mountain whitefish and rainbow trout showed better success at classifying sites having the species rather than sites without the species. The discrepancy in these cases ranged from 10% for mountain whitefish to 55% for rainbow trout.

Likelihood-ratio tests indicated that addition of field-based variables in the logistic models significantly improved model fit for 9 of the 14 species. Corresponding increases in ρ^2-values ranged from 0.061 for peamouths to 0.425 for prickly sculpin. The best map-and-field-based models showed an average improvement of 0.136 ($\pm$0.106 SD) in ρ^2-values versus the corresponding map-based models for these 9 species. Addition of bankfull width to the map-based variables generated the best models for chiselmouths, leopard dace, mountain whitefish, and redside shiners. Inclusion of both maximum temperature and bankfull width variables generated the best models for chinook salmon, largescale suckers, and peamouths. Addition of maximum stream temperature alone generated the best model for prickly sculpin. Logistic regression models for burbot, longnose dace, and rainbow trout did not improve with the incorporation of independent field variables. Inclusion of field variables failed to generate a significant regression for longnose sucker.

Model improvement following inclusion of field variables was not as apparent based on overall classification (Table 3). Although classification rates did improve considerably for prickly sculpin (+15%) and redside shiners (+7%) with inclusion of the best field variable predictors, the other species showed only a minimal improvement (0–4%) or reduced rates (i.e., chinook salmon = −4%, leopard dace = −5%). The average ($\pm$SD) improvement in classification rates across the nine species was only 2.8% ($\pm$5.6%) relative to the models using only map-based variables.

Models developed in the Blackwater drainage poorly classified seven species in the Similkameen drainage that were also found in the Blackwater drainage (i.e., species common to both drainages). Using the Blackwater map-based models, overall classification rates ranged from only 48% for mountain whitefish to 72% for northern squawfish, largescale suckers, and redside shiners; using the map-and-field-based models, rates ranged from 53% for mountain whitefish to 73% for largescale suckers and northern squawfish (Table 4). Average ($\pm$SD) overall classification rates for sites in the Similkameen drainage were 65.9% ($\pm$8.5%) for the map-based models and 67.3% ($\pm$6.9%) for the map-and-field-based models.

Logistic regression models developed for these seven species using habitat information derived solely from the Similkameen drainage greatly improved classification rates within this drainage. Using indigenous predictor variables for the Similkameen drainage, analogous to those from the Blackwater models, produced overall classification rates for map-based models ranging from 78% for rainbow trout to 93% for northern squawfish and prickly sculpin; rates for map-and-field-based models improved to 98% for prickly sculpin and largescale suckers (Table 4). The average ($\pm$SD) overall classification rates for sites in the Similkameen drainage using these unique regressions were 86.1% ($\pm$6.2%) for the map-based models and 88.1% ($\pm$8.2%) for the map-and-field based models.

Habitat conditions at stream sites within the Blackwater drainage appeared to broadly overlap with those found at sites in the Similkameen drainage (Table 5). However, sites in the Similkameen drainage were, on average, significantly colder,

TABLE 5.—Means (ranges) of habitat variables measured at stream sites in the Blackwater River drainage in 1996 and 1997 and in the Similkameen River drainage in 1995. Probability values of the Mann–Whitney *U*-test are given.

Habitat variable	Blackwater sites (N = 48)	Similkameen sites (N = 46)	P
Drainage area (km^2)	1,495 (13–11,830)	1,133 (2–7,808)	0.118
Watershed gradient (%)	1.6 (0.3–5.7)	3.0 (0.4–20.3)	0.042
Reach gradient (%)	1.1 (0.1–3.8)	3.3 (0.1–11.7)	0.001
Bank-full width (m)	19.4 (1.9–100.2)	31.6 (2.5–251.2)	0.500
Maximum depth (m)	1.06 (0.23–3.50)	0.91 (0.21–3.90)	0.242
Site elevation (m)	910 (750–1,170)	790 (150–1,350)	0.003
Maximum temperature (°C)[a]	20.1 (15.8–22.9)	12.8 (7.0–19.0)	0.000

[a] Maximum temperature for Blackwater drainage sites is based on annual temperature-logger readings; maximum temperature for Similkameen sites is based on hand-held thermometer readings in late summer.

steeper and lower in elevation than sites in the Blackwater drainage (Table 5).

Combining the fish present–absent and map-based habitat data sets from both drainages generated significant logistic regression models for each of the seven fish species common to both drainages (Table 6). Rho-square values for these combined models were considerably lower than the respective species models developed with drainage-specific data, except for longnose dace. Inclusion of a new categorical variable representing ‘‘drainage’’ (Blackwater or Similkameen) improved the species models significantly for five of the seven species, and the ‘‘drainage’’ predictor itself was either significant or approached significance in six out of seven regressions. However, classifications were still poorer than models specific to individual drainages.

Discussion

Logistic regression models can be used to identify specific habitat features that are associated with species presence and to predict the likelihood of occurrence of individual species. Watson and Hillman (1997) suggested that basin features such as watershed size, geology, and valley type provide the flow regimes and instream channel conditions (channel size, substrate, etc.) that different species of fish require for rearing, spawning, and migration. Individual drainages should therefore exhibit a specific and predictable array of large-scale physical characteristics that individual fish species require. Earlier applications of logistic models indicate that fish distributions can be successfully explained and predicted based on these macrohabitat variables. Logistic regressions incorporating

TABLE 6.—Logistic regression models for fish species found in both the Blackwater River and Similkameen River drainages using their combined map-based habitat data (N = 94). Model significance (P), degrees of freedom, and McFadden's ρ^2 is given for each species model. The significance (P) of the likelihood-ratio test (G-statistic) is given for evaluating improvement in each species models by including an artificial variable for ‘‘drainage’’ (Blackwater River or Similkameen River), as is the significance (P) of the individual ‘‘drainage’’ predictor within the species regression.

Species	Model input[a]	df	Model P-value	ρ^2	P-value of G-statistic	P-value of drainage category
Prickly sculpin	Map data	4	0.000	0.352		
	Map data plus drainage	5	0.000	0.401	0.032	0.047
Largescale sucker	Map data	4	0.000	0.219		
	Map data plus drainage	5	0.000	0.261	0.050	0.077
Longnose dace	Map data	4	0.000	0.308		
	Map data plus drainage	5	0.000	0.309	0.806	0.819
Mountain whitefish	Map data	4	0.000	0.265		
	Map data plus drainage	5	0.000	0.318	0.010	0.017
Northern squawfish	Map data	4	0.000	0.275		
	Map data plus drainage	5	0.000	0.494	0.000	0.003
Rainbow trout	Map data	4	0.009	0.147		
	Map data plus drainage	5	0.004	0.193	0.040	0.044
Redside shiner	Map data	4	0.000	0.205		
	Map data plus drainage	5	0.000	0.235	0.062	0.076

[a] Map data for each regression includes drainage area, watershed gradient, reach gradient, and elevation.

only channel slope and elevation, for example, correctly classified 87% of cutthroat trout sites in western Wyoming (Kruse et al. 1997). Watson and Hillman (1997) also produced a significant logistic model predicting presence or absence of bull trout at stream sites in Washington, Idaho, and Montana; they concluded that watershed-scale physical processes and spatial landscape patterns were the most important factors determining the species' distribution throughout the western United States.

In this study we developed logistic regression models for predicting fish species distributions in the Blackwater drainage based solely on a limited set of large-scale macrohabitat variables. High overall classification rates for many species models indicate that large-scale macrohabitat variables can be used for predicting the probability of occurrence of individual fish species at particular sites during summer low flows. The ability to derive significant logistic regression models is probably better for specialist species with narrow habitat requirements. Failure to generate successful predictive regressions for certain fish species (e.g., longnose sucker) suggests that it may be more difficult to develop incidence functions for highly generalist species. This could possibly be rectified by incorporation of additional variables, thereby more clearly defining the habitat niche of the species. This may require the addition of habitat variables that cannot be estimated at a macrohabitat scale.

Most fish species models were significantly improved by inclusion of field-based variables into the regressions. However, improvements in predictive success over strictly map-based models appeared to be small, the average increase being only 1.8% in site classifications across species. Although habitat conditions at a smaller scale (e.g., substrate, percent pool, large woody debris abundance, etc.) may be important determinants of fish abundance (Bowlby and Roff 1986: Jowett 1990; Watson and Hillman 1997), they may be of limited importance in determining presence or absence. Our results suggest that logistic models based solely on large-scale, map-derived variables may be sufficient for purposes of predicting fish species distributions within watersheds and that little may be gained by collecting field-based information for prediction purposes.

Our attempt to apply the Blackwater models outside the geographic boundaries of the Blackwater drainage was not successful (i.e., classification success at sites in the Similkameen drainage was low). This indicates that the relationship between large-scale habitat features and fish distributions may vary across broad land classes. Regional differences in landscape morphology are probably too great to allow wide use of the particular regression coefficients generated from the Blackwater sites. However, species regressions generated specifically for the Similkameen drainage produced much greater classification rates when applied to these same sites. This indicates that the selected macrohabitat variables have appreciable predictive power across drainages, but may require reestimation of regression coefficients for each drainage. Alternatively, we determined that combining fish and habitat data from drainages produced significant but less accurate models for species common to both drainages; model classification was improved by incorporating a simple categorical variable describing the respective drainages. This approach could be most useful for rarer species, which might require sampling from multiple drainages to obtain sufficient occurrence data to generate models.

Although we modeled fish presence as a function of distribution at summer base flow, fish may vary their habitat use seasonally (Shlosser 1982, 1985; Ross et al. 1985; Angermeir and Schlosser 1989), and additional models may be needed to predict spawning or overwintering distributions. Logistic models should also be refined to include factors such as the presence of downstream barriers (Kruse et al. 1997) that may influence species distributions. Age-specific regressions may also prove useful because habitat use and associated distributions may vary greatly between juveniles and adults of some species (Northcote 1997). Finally, biotic interactions such as competition and predation will also influence fish distributions (Watson and Hillman 1997), and the presence or relative densities of other fish species could be incorporated as independent variables in logistic regression.

Most studies of the effects of land-use practices on fish populations have been focused at relatively small spatial scales, usually confined to small- to moderate-sized stream reaches. However, to understand the effects of land-use practices on fish communities, a broader spatial context, such as watershed basin or subbasin, may be required. Watson and Hillman (1997) suggested that watershed analysis nested within a hierarchical framework of physical processes and management standards is the best choice for protecting fish biodiversity and maintaining the habitat of sensitive fish species. As part of the longer-term goal of integrated watershed management in British Columbia, the Blackwater and Similkameen drainages were selected as test sites for the development of predictive fish distribution models. The statistical models generated, using only macrohabitat variables derived from large-scale maps, appear to adequately explain the individual distributions of most fish species within these drainages. Although inclusion of field-based information generally produced better species distribution models, the predictive abilities of the models were only marginally improved. Further refinement of regression parameters could lead to the development of predictive models, which, based strictly on remotely derived variables, could map out provincial fish distributions based on probability of occurrence. Regional-scale models, based predominately on geomorphic variables and the habitat use of individual species, can be incorporated into geographic information systems, thereby allowing fisheries managers to assess and anticipate potential land-use impacts on fish populations. Although map-based models can enhance our understanding of fish distributions, it must be noted that classification success is never perfect, and fish absence or presence must be verified in the field. With this is mind, fish distribution models can provide a powerful diagnostic and predictive tool for improving watershed planning and management practices.

Acknowledgments

This project was conducted with funding from the Forest Renewal B.C and the Habitat Conservation Trust Fund. Our appreciation goes to Marianne Heise, John Pritchard, Anne Salomon, Bev Wicks, Pier Van Dishoek, and Michelle Roberge for enthusiastic and able assistance in the field. Don McPhail and Gordon Haas provided useful advice and Maurice Lirette provided essential local support and expertise. We would also like to thank Pete and Cindy Carter for their unrivalled hospitality.

References

Angermeier, P. L., and J. R. Karr. 1983. Fish communities along environmental gradients in a system of tropical streams. Environmental Biology of Fishes 9:117–135.

Angermeier, P. L., and I. J. Schlosser. 1989. Species-area relationships for stream fishes. Ecology 70: 1450–1462.

Bjornn, T. C., and D. W. Reiser. 1991. Habitat requirements of salmonids in streams. Pages 83–138 *in* W. R. Meehan, editor. Influences of forest and rangeland management on salmonid fishes and their habitats. American Fisheries Society, Special Publication 19, Bethesda, Maryland.

Bowlby, J. N., and J. C. Roff. 1986. Trout biomass and habitat relationships in southern Ontario streams. Transactions of the American Fisheries Society 115: 503–514.

Cannings, S. G., and J. Ptolemy. 1998. Rare freshwater fish of British Columbia. Province of British Columbia, Ministry of Environment, Lands and Parks, Victoria.

Eadie, J. M., and A. Keast. 1982. Do goldeneye and perch compete for food? Oecologia 55:225–230.

Fausch, K. D., and T. G. Northcote. 1992. Large woody debris and salmonid habitat in a small coastal British Columbia stream. Canadian Journal of Fisheries and Aquatic Sciences 49:682–693.

Gorman, O. T., and J. W. Karr. 1978. Habitat structure and stream fish communities. Ecology 59:507–515.

Griffith, R. P. 1991. Blackwater (West Road) River assessment of existing and potential resident rainbow trout production downstream of Kluskoil Lake. Report (CA-914) of R. P. Griffith and Associates to British Columbia Environment, British Columbia Regional Fisheries, Williams Lake.

Hubert, W. A., and S. J. Kozel. 1993. Quantitative relations of physical habitat features to channel slope and discharge in unaltered mountain streams. Journal of Freshwater Ecology 8:177–183.

Johnston, N. T., and P. A. Slaney. 1996. Fish habitat assessment procedures. Province of British Columbia, Ministry of Environment, Lands and Parks and Ministry of Forests, Watershed Restoration Technical Circular 8, Vancouver.

Jowett, I. A. 1990. Factors relating to the distribution and abundance of brown and rainbow trout in New Zealand clear-water rivers. New Zealand Journal of Marine and Freshwater Research 24:429–440.

Kruse, C. G., W. A. Hubert, and F. J. Rahel. 1997. Geomorphic influences on the distribution of Yellowstone cutthroat trout in Absaroka Mountains, Wyoming. Transactions of the American Fisheries Society 126:418–427.

Lanka, R. P., W. A. Hubert, and T. A. Wesche. 1987. Relations of geomorphology to stream habitat and trout standing stock in small Rocky Mountain streams. Transactions of the American Fisheries Society 116:21–28.

Lobb, M. D., and D. J. Orth. 1991. Habitat use by an assemblage of fish in a large warmwater stream. Transactions of the American Fisheries Society 120: 65–78.

McPhail, J. D., and R. Carveth. 1992. A foundation for conservation: the nature and origin of the freshwater fish fauna of British Columbia. University of British Columbia, Department of Zoology, Fish Museum, Vancouver.

Meehan, W. R., and T. C. Bjornn. 1991. Salmonid distributions and life histories. Pages 47–82 *in* W. R. Meehan, editor. Influences of forest and rangeland management on salmonid fishes and their habitats. American Fisheries Society, Special Publication 19, Bethesda, Maryland.

Moyle, B. M., and D. M. Baltz. 1985. Microhabitat use by an assemblage of California stream fishes: developing criteria for instream flow determinations. Transactions of the American Fisheries Society 114: 695–704.

Nelson, R. L., W. S. Platts, D. P. Larsen, and S. E. Jensen. 1992. Trout distribution and habitat in relation to geology and geomorphology in the North Fork Humboldt River drainage, northeastern Nevada. Transactions of the American Fisheries Society 121: 405–426.

Newbury, R. W., and M. C. Gaboury. 1994. Stream analysis and fish habitat design; a field manual, 2nd edition. Newbury Hydraulics, Gibsons, British Columbia.

Nickelson, T. E., J. D. Rodgers, S. L. Johnson, and M. F. Solazzi. 1992. Seasonal changes in habitat use juvenile coho salmon (*Oncorhynchus kisutch*) in Oregon coastal streams. Canadian Journal of Fisheries and Aquatic Sciences 49:790–794.

Northcote, T. G. 1997. Potamodromy in Salmonidae—living and moving in the fast lane. North American Journal of Fisheries Management 17:1029–1045.

Parkinson, E. P., A. Tautz, G. Haas, and J. Rosenfeld. 1996. Forestry related research needs for managing B. C.'s freshwater fish diversity. Province of British Columbia, Ministry of Environment, Lands and Parks, Fisheries Project Report RD 61, Victoria.

Press, S. J., and S. Wilson. 1978. Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association 73:699–705.

Rieman, B. E., and J. D. McIntyre. 1995. Occurrence of bull trout in naturally fragmented habitat patches of various size. Transactions of the American Fisheries Society 124:285–296.

Rosenfeld, J. 1996. Fish distribution, diversity, and habitat use in the Similkameen watershed. Province of British Columbia, Ministry of Environment, Lands and Parks, Fisheries Project Report 52, Victoria.

Ross, S. T., W. J. Matthews, and A. H. Echelle. 1985. Persistence of stream fish assemblages: effects of environmental change. American Naturalist 126: 24–40.

Schlosser, I. J. 1982. Fish community structure and function along two habitat gradients in a headwater stream. Ecological Monographs 52:395–414.

Schlosser, I. J. 1985. Flow regime, juvenile abundance, and the assemblage structure of stream fishes. Ecology 66:1484–1490.

Steinberg, D., and P. Colla. 1991. LOGIT: a supplementary module for SYSTAT. SYSTAT, Evanston, Illinois.

Tabachnick, B. G., and S. Fidell. 1996. Using multivariate statistics, 3rd edition. HarperCollins, New York.

Water Survey of Canada. 1989. Historic streamflow summary to 1988: British Columbia. Environment Canada, Inland Waters Directorate, Ottawa.

Watson, G. W., and T. W. Hillman. 1997. Factors affecting the distribution and abundance of bull trout: an investigation at hierarchical scales. North American Journal of Fisheries Management 17:237–252.