USDA Forest Service
General Technical Report RM-87

August 1981

The Use of Multivariate Statistics in Studies of Wildlife Habitat[1]

Edited by

David E. Capen
University of Vermont

The workshop at which the papers included in this report
were presented was jointly sponsored by

School of Natural Resources, University of Vermont

U.S. Fish and Wildlife Service

USDA Forest Service

[1]This report was published by the Rocky Mountain Forest and Range Experiment Station. Supervision was provided by Thomas W. Hoekstra, project leader for the National Resource Analysis Techniques Project of the Resource Evaluation Techniques Program. Station headquarters is in Fort Collins, in cooperation with Colorado State University.

AN APPLICATION OF FACTOR ANALYSIS IN AN AQUATIC HABITAT STUDY[1]

T.J. Harshbarger[2], and H. Bhattacharyya[3]

Abstract.--In five small, high-gradient trout streams in western North Carolina, 18 cover variables were related to standing crop biomass of wild brook trout (Salvelinus fontinalis), rainbow trout (Salmo gairdneri) and brown trout (Salmo trutta) in randomly selected stream sections. Factor analysis of the data set showed that only a small number of factors or variables was needed to explain relations between variables in the observed set. Key cover factors were area in debris; turbulent water; vegetation, both in and over stream; and overhanging banks. Resolutions obtained were used in stepwise regressions to explore relationships between standing crop of trout and age of fish. Regressions containing factors as independent variables explained less variation in fish standing crop than did regressions containing equal numbers of original habitat attributes as independent variables.

Key words: Aquatic habitat; cover; factor analysis; multivariate analysis; regression analysis; stream fish; trout.

INTRODUCTION

The presence of a self-sustaining population of fish usually indicates compatibility between the aquatic envrionment and the ecological requirements of the fish. Wild trout are excellent indicators of current environmental conditions and their population density in streams reflects their level of compatibility with a highly integrated chemical, physical and biological situation. However, simply realizing that the trout population reflects its envrionment is not particularly informative to the resource manager. Explicit relationships between species and their environments are needed to assess the actual and potential capabilities of the habitat.

Functional and correlative approaches have been used to study factors influencing the distribution and abundance of a species. The functional approach is used when factors are known to influence certain attributes of the species. Correlative procedures are best suited for exploratory studies, where the relationship between a species and its environment are unknown. This procedure provides little information about causality, but helps the researcher to make inference for rigorous testing. Many variables can potentially influence the distribution and density of wild trout in a stream. Often the choice of parameters to measure and analyze is difficult because environmental variables in lotic waters are typically correlated and confounded with one another (Reid 1961).

Attempts to correlate single and multiple variables to trout populations in streams have met with varying degrees of success. Boussu (1954), Saunders and Smith (1962), and Wickham (1967) investigated the relationship of a trout

[1]Paper presented at The use of multivariate statistics in studies of wildlife habitat: a workshop, April 23-25, 1980, Burlington, Vt.

[2]Aquatic Ecologist, Southeastern Forest Experiment Station, Asheville, NC 28806.

[3]Mathematical Statistician, Southeastern Forest Experiment Station, Research Triangle Park, NC 27709.

population to cover. In two studies (Schuck 1943, Hunt 1969), water depth strongly influenced trout densities in streams, and current velocity was similarly implicated in another study (Lewis 1969). Few investigators, notably Lewis (1969), Stewart (1970), Platts (1974), and Binns and Eiserman (1979), have studied the simultaneous effect of several environmental variables on stream trout populations using multiple regression and correlation analysis.

The present study was designed to look at the relationships between various habitat parameters and trout. Factor analysis was used to delineate and examine a group of environmental variables that seemed important to trout, those providing cover or shelter. Regression analysis was used to examine relationships between the environmental factors produced and trout.

STUDY SITES AND METHODS

We selected five small, high-gradient trout streams in western North Carolina. In each, we inventoried and measured 18 variables providing fish cover, in 20 randomly selected 30m stream sections. Each section was surveyed along line transects established every 3m across the stream, perpendicular to its center line. All physical structures providing shelter or concealment for trout were located and their cross-sectional areas parallel to the air-water interface recorded. Structures included rocks or ledges which afforded cover to fish, undercut banks, aquatic vegetation, and logs and brush in the stream. Other cover situations such as water with sufficient surface turbulance to prevent visibility of the stream bottom were also measured and recorded.

Brush and loosely compacted debris in the stream and streamside vegetation overhanging the water surface were estimated ocularly. Cover afforded by brush and debris was recorded as the surface occupied by solid material and expressed as a percentage of the water surface it covered. Cover provided by overhanging bank vegetation was expressed in two variables: as the percentage of stream covered by vegetation between the water surface and 1.0m above, and the percentage between 1.0m and 2.0m above the surface.

Standing crop biomass of wild trout was estimated by depletion analysis electrofishing. All fish were weighed, measured, and returned to the midpoint of the section under study.

Factor Analysis

Starting with the observed correlation matrix on the 18 cover variables, we used SAS FACTOR programs for principal axis, maximum likelihood, and iterated principal axis factoring (SAS 1979). Using the principal axis method and keeping only those factors corresponding to eigenvalues greater than one, six factors were retained. To make the factors more conceptually meaningful, several rotational methods were used, including varimax, quartimax, equimax and promax (oblique) procedures (SAS 1979).

All factor and rotation procedures used on the 18 cover variables produced essentially the same information about underlying structure. However, the iterated principal axis and oblique rotational procedures are intuitively appealing. These procedures address communalities directly and recognize that some factors may very likely be correlated. As such only the interated principal axis solution using an oblique rotation of the cover related factors will be presented in this paper.

Regression Analysis

Relationships between standing crop of trout and the factors obtained from oblique rotation of the iterated principal axis solution were explored using SAS STEPWISE regression procedures and the maximum R^2 improvement technique developed by Goodnight (SAS 1979). Relationships between the habitat variables themselves and standing crop of trout were similarly explored to determine if factors entered the stepwise model in about the same order as the variable they contained. Coefficients of determination were compared to assess performance of models containing factors and variables.

RESULTS

The six factors combined the 18 cover variables into groups which generally reflected meaningful patterns in relation to the stream environment. Factors were named after the variable or variables producing greatest correlation. Factor loadings for the variables are shown in table 1. Factors had zero or close to zero projections on most variables, very few immediate loadings, and two or three high associations.

Factor 1 is a measure of debris with high loadings for number of logs and the surface areas, parallel to the stream, in logs and in brush. Factor 2 is a measure of side stream cover and is strongly negatively correlated with area in and percent cover of vegetation trailing in the water surface. The third factor expresses the percentages of cover provided by overstream vegetation 0-1m and 1-2m above the stream. Factor 4 is highly correlated with area in turbulent water and to a lesser degree with the number of units of turbulent water. This factor is also correlated with total cover and can be considered a general cover factor. Rock area and number load heavily on factor 5, and the sixth factor is highly correlated with area in overstream vegetation in the 1- to 2-m zone.

Factor values for each of the 100 stream section observations were obtained from the scoring coefficient matrix. Trout, regardless of

Table 1. Variables associated with cover factors.

Variables	Factor loadings 1	2	3	4	5	6
Ledge area (1)	0.063	0.241	0.030	0.400	-0.288	0.106
Rock						
Number (2)	-0.075	-0.047	-0.009	-0.096	0.850	0.214
Area (3)	0.099	0.194	0.114	0.109	0.646	-0.191
Turbulent water						
Number (4)	-0.108	0.304	-0.075	0.491	0.007	0.156
Area (5)	-0.244	-0.054	-0.059	0.936	-0.115	-0.014
Logs						
Number (6)	0.913	-0.049	-0.007	-0.108	-0.146	-0.003
Area (7)	0.969	0.003	-0.060	-0.089	-0.019	0.020
Bank area (8)	0.080	-0.539	-0.039	-0.125	-0.119	-0.080
Other area (9)	0.045	0.002	0.345	-0.051	0.112	-0.090
Brush						
% Cover (10)	0.102	-0.150	0.282	0.062	0.033	-0.012
Area (11)	0.824	0.076	0.067	-0.006	0.084	0.076
Side-stream vegetation						
% Cover (12)	-0.116	-0.772	0.048	-0.087	0.082	0.198
Area (13)	0.001	-0.712	0.006	0.012	0.011	0.015
Over-stream vegetation (0-1m)						
% Cover (14)	-0.045	-0.042	0.896	0.034	-0.109	-0.201
Area (15)	0.174	0.217	0.111	0.231	-0.333	0.363
Over-stream vegetation (1-2m)						
% Cover (16)	-0.032	0.089	0.844	-0.110	0.039	0.293
Area (17)	0.048	-0.050	-0.062	0.046	0.094	0.999
Total cover area (18)	0.365	-0.105	0.062	0.775	0.225	-0.076

species, in each stream section were segregated into the four age classes represented in each of the five streams under study. The relationship between factor values and trout standing crop for each age group was examined using the stepwise maximum R^2 regression technique. Likewise, the relationship between the original 18 cover attributes and trout standing crop was similarly examined.

The coefficient of determination (R^2) corresponding to the six factors was lower than that corresponding to the original 18 variables for each age class of fish (table 2). Further, the best one-attribute, two-attribute, etc., models obtained by the stepwise procedure also produced higher R^2 values when the original variables were used than when the deduced factors were used. Six-factor models produced coefficients (R^2) which ranged from 0.09 for the standing crop of the young-of-the-year trout to 0.53 for trout in age group II. Regressions on the original variables produced coefficients between 0.31 and 0.71 for 18 variable models and 0.26 and 0.66 for models containing six variables.

For young-of-the-year trout a single variable, rock area, produced an R^2 value (0.10) equivalent to that obtained from the model containing all six factors. In age group I, two habitat variables, numbers of rocks and total cover, produced a coefficient (0.18) as large as the six-factor model. A coefficient similar to that produced by the six-factor model was obtained for trout in age group II with three variables (R^2=0.56); number of rocks, percent cover provided

Table 2. Order in which the habitat attributes shown in table 1 entered stepwise regressions and the coefficient of determination they produced for each age group of trout.

Age group	Step 1 Attributes (R^2)	Step 2 Attributes (R^2)	Step 3 Attributes (R^2)	Step 4 Attributes (R^2)	Step 5 Attributes (R^2)	Step 6 Attributes (R^2)
Group 0						
Factors	6 (0.03)	4,6 (0.05)	3,4,6 (0.07)	2,3,4,6 (0.08)	1,2,3,4,6 (0.09)	1,2,3,4,5,6 (0.09)
Variables	3 (0.10)	2,3 (0.16)	2,3,12 (0.23)	3,12,15,27 (0.29)	2,3,12 15,17 (0.31)	2,3,10,12, 15,17 (0.32)
Group I						
Factors	5 (0.08)	4,5 (0.12)	1,4,5 (0.14)	1,2,4,5 (0.15)	1,2,3,4,5 (0.15)	1,2,3,4,5,6 (0.15)
Variables	2 (0.14)	2,18 (0.18)	1,2,10 (0.21)	1,2,3,10 (0.23)	1,2,3,4,12 (0.25)	1,2,3,4,10,12 (0.26)
Group II						
Factors	6 (0.17)	5,6 (0.40)	4,5,6 (0.46)	2,4,5,6 (0.49)	1,2,4,5,6 (0.52)	1,2,3,4,5,6 (0.53)
Variables	2 (0.31	2,3 (0.44)	2,10,17 (0.55)	2,10,15,17 (0.61)	2,10,12,15,17 (0.64)	2,10,12,15, 17,18 (0.66)
Group III						
Factors	6 (0.16)	5,6 (0.27)	2,5,6 (0.31)	1,2,5,6 (0.35)	1,2,4,5,6 (0.36)	1,2,3,4,5,6 (0.37)
Variables	17 (0.23)	15,17 (0.29)	12,15,17 (0.35)	4,12,15,17 (0.38)	2,4,12,15,17 (0.40)	2,4,10,12, 15,17 (0.42)

by instream brush, and area in overstream cover between 1 and 2 m; and for trout in age group III with four variables (R^2=0.34); number of pockets of turbulent water, percent cover of side-stream vegetation, area in overstream cover to 1 m, and area in overstream cover to 2 m.

Factor 5 (rocks) and factor 6 (area in overstream cover 2 m and above) and their equivalent habitat attributes were, in most instances, the first parameters to enter stepwise regressions containing factors or variables. Factor 2 (overstream cover) and factor 4 (general cover) entered models intermediately, and factor 1 (debris) and 3 (percent overstream cover) entered models last. Area in and percent cover provided by side-stream vegetation consistently entered models intermediately. Other habitat attributes either entered models intermediately or late in the stepwise procedure.

DISCUSSION

Factor analysis was useful in defining the underlying structure of the cover portion of the habitat for trout. Although 18 variables were singled out and measured, cover consists essentially of a six-dimensional space characterized by six factors: debris, side-stream cover, percent of overstream vegetation, turbulent water, rock area, and area of overstream vegetation in the 1- to 2-m zone.

In examining the relationships between standing crop of fish and the habitat attributes by stepwise regression methods, a higher coefficient of determination (R^2) was obtained by using the 18 original variables than by using the six derived factors. Moreover, it was found that the best one-attribute, two-attribute, etc., models also resulted in higher R^2 values when original variables were used. This indicates that when the original variable measurements are available, there is no reason to form regression models based on derived factors. With the exception of 2-year-old trout, no R^2 value between trout and the set of variables exceeded 0.5, indicating that there is a substantial portion of variability in fish biomass that is not explained by the measured variables and hence also not explained by the derived factors. Part of this variability may be accounted for by water or flow related variables. The results of a combined analysis of water and cover variables will be reported at a later date.

LITERATURE CITED

Boussu, M.F. 1954. Relationship between trout populations and cover on a small stream. Journal of Wildlife Management 18:229-239.

Binns, N.A., and F.M. Eiserman. 1979. Quantification of fluvial trout habitat in Wyoming. Transactions of the American Fisheries Society 108:215-228.

Hunt, R.L. 1969. Effects of habitat alteration on production, standing crops, and yield of brook trout in Lawrence Creek, Wisconsin. p. 281-312. In T.G. Northcote, editor. Symposium on salmon and trout in streams. H.R. MacMillan Lectures in Fisheries, Vancouver, Canada.

Lewis, S.L. 1969. Physical factors influencing fish populations in pools of a trout stream. Transactions of the American Fisheries Society 98:14-19.

Platts, W.S. 1974. Geomorphic and aquatic conditions influencing salmonids and stream classification--with application to ecosystem classification. USDA Forest Service, Intermountain Forest and Range Experiment Station, Boise, Idaho.

Reid, G.K. 1961. The ecology of inland waters and estuaries. 375 p. Reinhold, New York, N.Y.

SAS. 1979. Users guide. 494 p. SAS Institute, Inc., Raleigh, N.C.

Saunders, J.W., and M.W. Smith. 1962. Physical alteration of stream habitat to improve brook trout production. Transactions of the American Fisheries Society 91:185-188.

Schuck, H.A. 1943. Survival, population density, growth, and movement of the wild brown trout in Crystal Creek. Transactions of the American Fisheries Society 73:209-230.

Stewart, P.A. 1970. Physical factors influencing trout density in a small stream. Ph.D. Dissertation, Colorado State University, Fort Collins, Colo.

Wickham, M.G. 1967. Physical microhabitat of trout. M.S. Thesis, Colorado State University, Fort Collins, Colo.

DISCUSSION

TERRY LARSON: Most papers given here which involved multiple regression analysis used stepwise techniques. Why did you not use an all possible subsets technique like BMDP-9R? This program is not costly to run and will pick up subsets that stepwise techniques miss.

HELEN BHATTACHARYYA: The SAS STEPWISE procedure with MAXR option was used. You are quite right that all possible regression (SAS RSQUARE procedure) may pick out combinations not covered by stepwise. However, STEPWISE/MXR is almost as good (see SAS writeup) and has the advantage of printing all other regression statistics, slope, intercept, mean squares, etc., besides just the R^2 value.