THE UNIVERSITY OF MANITOBA

SPECIES DIVERSITY AND WITHIN-STREAM DISTRIBUTION OF

STREAM DWELLING FISHES FROM WESTERN MANITOBA

BY

HUGH VALIANT

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF ZOOLOGY

WINNIPEG, MANITOBA

"SPECIES DIVERSITY AND WITHIN-STREAM DISTRIBUTION OF STREAM DWELLING FISHES FROM WESTERN MANITOBA"
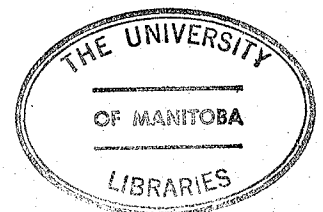
by

HUGH VALIANT

A dissertation submitted to the Faculty of Graduate Studies of
the University of Manitoba in partial fulfillment of the requirements
of the degree of

MASTER OF SCIENCE

© 1975

ABSTRACT

Abundance of fishes was determined at stations located in several western Manitoba streams. By use of a multivariate classification procedure, groups of species were identified which were similar in their distribution among arbitrarily defined habitat types. The largest group of species was associated with areas of low current speed, but within high gradient sections of the streams. Species diversity (defined as the number of fish species present) was related to structural features of the environment. By an analysis of variance technique, it was concluded that stream gradient and current speed explained most of the variation in diversity. When a predator (Esox lucius) was present, diversity and abundance of the remaining species were both reduced.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

Table of Contents (continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

INTRODUCTION

This study attempts to examine the relationship between species diversity of stream fishes and certain structural features of their environment.

It is generally believed that the diversity of biological communities is related to environmental factors, but the nature of this relationship appears to vary considerably between different geographic areas or taxonomic groups. No model has been suggested which fits more than a small subset of the different community types. Even for a group as relatively homogeneous as temperate stream fishes, Hynes (1970) has suggested that the relationship between diversity and environmental factors depends a great deal on the identity of the component species. Therefore, in this study the discussion of species diversity was preceded by a description of the within-stream distributions of individual species. Also, some precision within any given stream was sacrificed in order to obtain an estimate of "among streams" variability, and hence an estimate of the consistency of the results.

THE STUDY AREA

A total of six streams in the escarpment region of western Manitoba were sampled during the summers of 1973 and 1974 (Fig. 1). These streams are small, generally less than four metres in width and one metre in depth. With one exception, it is possible to divide each stream into two sections based on gradient: a low gradient section in the upstream areas, and a higher gradient section closer to the

stream mouth. The low gradient sections are meandering and silt-bottomed, often with considerable aquatic vegetation. These streams typically increase in gradient near their mouths as they enter the valley of a larger river. Plum Creek flows southeast into the Souris River at Souris, Manitoba. Oak Creek flows north into the Souris near Treesbank, Manitoba. The Cypress River flows north, entering the Assiniboine River east of Spruce Woods Park. The Pembina River flows east out of the Turtle Mountains. Neepawa Creek flows south-east out of the Riding Mountains and joins the Whitemud River, which flows into Lake Manitoba. The Garland River flows east out of the Duck Mountains into Lake Winnipegosis.

At first, only four of the streams were sampled, but the upper Pembina River was added in mid-summer of 1973 and the Cypress River was added during the summer of 1974 as a substitute for Plum Creek, which went dry the previous year.

Figure 1. Map of study area, showing locations of streams.

Lake
Winnipegosis

Lake
Manitoba

Garland   R

Dauphin
Lake

MANITOBA

50 Km

Neepawa Cr.

Whitemud R.

Assiniboine   R.

Plum Cr.

Oak Cr.

Cypress R.

Souris R.

Pembina R.
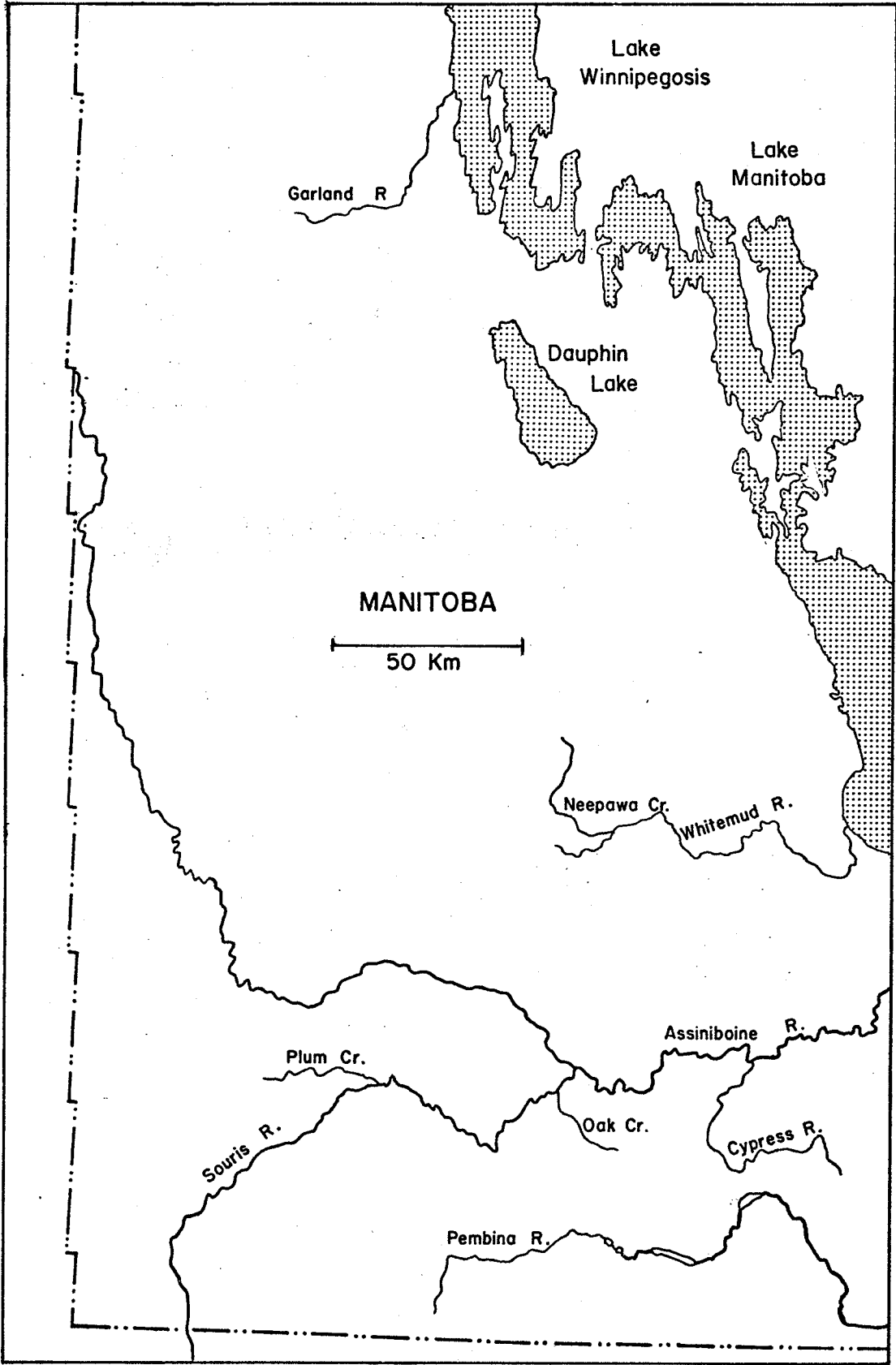
METHODS AND MATERIALS

SAMPLING

In 1973, twelve stations were chosen on each of the four original streams. These stations were allocated as follows: each stream was divided into two sections, based on gradient. Within each section, three shallow and three deep stations were chosen so that they consisted of one each of fine (silt or sand), medium (gravel) and coarse (rocky) substrates. Thus, each combination of two gradients x three substrates x two depths occurred once in each stream. Later, after some sampling had been done, it became clear that treating depth and substrate size as independent factors was somewhat unrealistic since, in running water, depth, substrate size and current speed are closely related. It was decided to combine the depth and substrate factors and to just call each station a riffle, channel or pool.

Each station was sampled in sequence to form one replicate. Four replicates were obtained, one in each of May, June, July and August. Replications are called "seasons" in the discussion of results, since their effect was one of seasonal changes. The nature of seasonal changes in species diversity was not known a priori, but all stations were sampled in the same order each time so as to keep the time between sampling periods nearly constant for each station. It was thought that this might reduce unexplainable variability.

In 1974, attention was confined to the high gradient sections of the streams, with only six stations (two channels, two riffles,

two pools) in each stream. This change was made because stations
in the low gradient areas were quite consistent in containing few
species and it was decided that continued sampling effort in those
areas would yield little new information.

MEASUREMENT OF DIVERSITY AND ASSOCIATED VARIABLES

In order to determine species diversity, each station was
enclosed using barrier nets and seined repeatedly until three
consecutive hauls yielded no fish. Fish taken from a station were
identified and number of each species counted. In 1974, all fish
taken were measured to the nearest millimeter fork length. Some
additional measurements were taken to obtain more information about
habitats occupied by individual species. Stream width was recorded
at two points, two meters apart. Associated with each width
measurement was a set of three depth measurements at 1/4, 1/2 and
3/4 of channel width. Water temperatures and time of day were
recorded at each station. In 1973, a measure of surface velocity
was obtained by timing a float over a fixed distance. In 1974, a
current meter became available and current speed was obtained at the
surface and next to the substrate at each point that a depth measure-
ment was taken. In 1973, water chemistry characteristics were
determined using a Hach kit, but since little variation was observed
from station to station, this was abandoned the next year.

METHODS OF ANALYSIS

A study of the factors affecting species diversity is basically
a study of the factors affecting niche overlap (Hutchinson 1959).

In Appendix 1, an argument is given for advantages of the species number, S, over more complex diversity indices, to the effect that other indices confound relative abundance and number of species. A certain loss of information is to be expected with any summary measure including S, however, and therefore it was decided to first consider the distribution of individual species in the hope that this would clarify the discussion of species diversity.

      1. <u>Distribution of species among habitats</u>: In the overall sampling design, habitats were divided into three types: riffles, channels and pools. This rather coarse separation of habitats was necessary because each habitat was sampled in four streams, in two gradient sections and at four times during the first summer, and the product of these factors was near the maximum of stations it was possible to consider. When considering the "habitats" factor by itself, however, a finer division of habitat types was possible, and so the riffles, channels and pools defined in the sampling design were arbitrarily reclassified to form six habitat types with depths and current speeds as in Table 1. Not all habitats could be separated on depth and current speed alone. Deep areas were divided into three types (4, 5, 6) which were similar in depth and current speed, but which differed in shape, in substrate type and in proximity to riffles. Pools below riffles (type 4) generally had gravel substrates with no silt deposited. Type 5 pools were variable in substrate type: usually gravel with some silting evident. They were discrete units in the sense that they were bordered by shallower

Table 1. Defined Habitat Types

	Riffle (1)	Fast Channel (2)	Moderate Channel (3)	Pool Below Riffle (4)	Pool (5)	Deep Slow Channel (6)
Current speed (cm./sec.)	.75-2.0	.3-1.0	.15-.5	0-.15	0-.1	0-.1
Depth (cm.)	5-25	8-25	25-50	35-100	35-100	60-110
Substrate	Rock	Sand, gravel	Gravel	Gravel	Variable	Silt

areas, although they were not immediately adjacent to riffles. Deep slow channels (type 6) were sections of longer channels with slow flowthrough rate and silt and clay substrates. Since substrate type and proximity to upstream riffles determines to some extent the nature of the invertebrate bottom fauna (Hynes 1970) and drift (Waters 1972) respectively, it was expected that differences in fish species composition might exist among these three habitats.

No attempt was made to estimate the distributions of species which were represented by fewer than twenty individuals over the two year period. The distribution of each species was then expressed as a set of six probabilities, of observing each species in each habitat type. The probability of observing species (i) in the (j)th habitat type was estimated as

$$P_{ij} = S_{ij} / N_{ij} \cdot T_i, \quad i = 1, 2, \ldots, 13; \quad j = 1, 2, \ldots, 6,$$

where S_{ij} = number of species (i) observed in all stations
of type (j),

N_{ij} = number of stations of type (j) in streams in
which species (i) was present,

$$T_i = \sum_j S_{ij} / N_{ij}.$$

In words, P_{ij} is the fraction of individuals belonging to species (i) which were found in habitat type (j); the number of habitats of each type available to species (i) was defined to be the number of stations of each type which occurred in the streams in which species (i) was present. P_{ij} is thus independent of both species (i)'s population size and identity of streams in which

species (i) was present. This allows comparison of species on a
basis of the proportion of each species observed in each habitat type
without regard to abundance or streams occupied. These probabilities
were determined from those stations for which length measurements
were made.

The thirteen species were then successively combined to form
groups which were similar in habitat selection, according to the
following procedure, modified from Orloci (1967): the within-group
dispersion of a group A, consisting of n species, was defined as

$$Q_A = \sum_j \sum_i (p_{ij} - \bar{p}_j)^2$$

where $\bar{p}_j$ is the average probability, for all species, of occurring in
habitat (j). (This is just the total squared distance of the points
in the group from their average, and is thus a measure of the group's
heterogeneity.) At any step in the procedure, two groups A and B
were joined to form a new group AB if

$$Q_{AB} - (Q_A + Q_B) \leq Q_{CD} - (Q_C + Q_D), \forall \ C, D \ \varepsilon \ U, \text{ where U is the}$$
set of all possible groups.

Thus, if two groups were fused at an early stage of the
grouping procedure, the two groups were considered similar in
habitat selection.

Only those stations at which all fish had been measured were
considered; several authors have indicated (Gibbons and Gee 1972;
Trautman 1957) that smaller fish of some species occupy a different
spatial niche than older, larger individuals. Gibbons and Gee (1972),
for example, demonstrated that fry of both Rhinichthys species occupied

areas of lower current speed than adults. Therefore, a measure of size for each fish was necessary in order to account for the presence of fish of different ages. Length measurements were made in August, 1973 and throughout 1974. To account for the possible differences in spatial niche between small and large individuals within species, an attempt was made to separate each species into two size classes based on length frequency plots. Where possible, the size classes conformed to fry and older age classes, respectively. However, blackside darters were insufficiently abundant to make a decision on this basis, and were divided arbitrarily into small and large size classes. White suckers, sand shiners and fathead minnows were not separated. White suckers spawn in small streams, but the offspring move into larger bodies of water before age one (Scott and Crossman 1973). Sand shiners were encountered only in those streams flowing into the Souris River, and in the Pembina River. They were uncommon in these small streams, but were the most abundant species present in a few seine hauls made in the Souris River. Scott and Crossman (1973) describe this species as preferring lakes and large rivers. Since individuals found in the small streams were quite uniform in size, and no fry were observed, no attempt was made to separate this species into size classes. The size frequency distribution of fathead minnows appeared to be unimodal, with little variability so this species was likewise not divided into size classes. The criteria for separation of size classes are given in Table 2.

The original intent was to treat each size class as a separate "species" and then repeat the above analysis, but computer

Table 2. Criteria for Separation of
          Size Classes

Species	Smaller size class if <
Etheostoma nigrum	39
Percina maculata	44
Notropis dorsalis	32
Notropis stramineus	--
Notropis cornutus	32
Semotilus atromaculatus	50
Rhinichthys cataractae	40
Rhinichthys atratulus	30
Catostomus commersoni	--
Pimephales promelas	--
Semotilus margarita	41
Chrosomus neogaeus	40
Culaea inconstans	32

space limitations (APL/360) made a modification to this approach necessary. The analysis was repeated, using probabilities calculated for the larger of the two size classes only, for the ten species which were divided. The remaining three species were included as before.

2. Species Diversity: To examine the effects of environmental variables on species diversity, data collected in 1973 were analyzed as an analysis of variance, with factors as defined in the sampling design. As in the usual ANOVA layout, interactions between factors were included as part of the model, but it was decided to include only the two-way interactions since higher order interactions are often difficult to interpret and would add considerably to complexity of the model.

Since diversity of species was observed in the same set of stations at four times, the values of diversity observed in one station at successive times are not independent; for example, there is a possibility that, if a given station gave a high yield at one time, it might have been expected to give a high yield in the next time period. Therefore, observations on the same station at different times do not constitute true replications: the correct error term for testing stream, gradient and habitat effects is a mean square based on replications of these treatments in space (this will be called "replications of sites" in subsequent discussion). This type of design is called a "split plot" (Snedecor and Cochran 1967).

One of the assumptions necessary for F-tests to be valid is that variances within the treatment combinations are the same

(homogeneity of residual variances). A plot of the residuals, or
the difference between the predicted and the observed Y values, versus
the predicted Y values, is given in Fig. 2, for Y = S, the number of
species. Since the numbers of species were all small integers, a
Y = $\sqrt{S + 1}$ transformation was applied to the data. This seemed to
improve the distribution of the residuals slightly (Fig. 3) and
also appeared to increase additivity somewhat, and so this transformation
was retained in the subsequent analysis.

Because of changes in the original layout, and because some
observations were missing due to other causes, such as stations going
dry, the design lost its original orthogonality. This means, in effect,
that the sum of squares attributable to treatments cannot be partitioned
into sums of squares due to individual factors, because of covariances
among the factors. As a result, the significance of each factor must
be tested individually (Armitage 1971) by calculating a separate
regression on all independent variables except those representing the
factor in question, and then determining the reduction in the sum of
squares due to treatments for this "reduced" model against the complete
model. Details on the model used and methods of calculation are given
in Appendix 2.

Figure 2. Plot of residuals versus expected Y-values, for Y = S.

RESIDUALS

EXPECTED V-VALUE, V= S

16
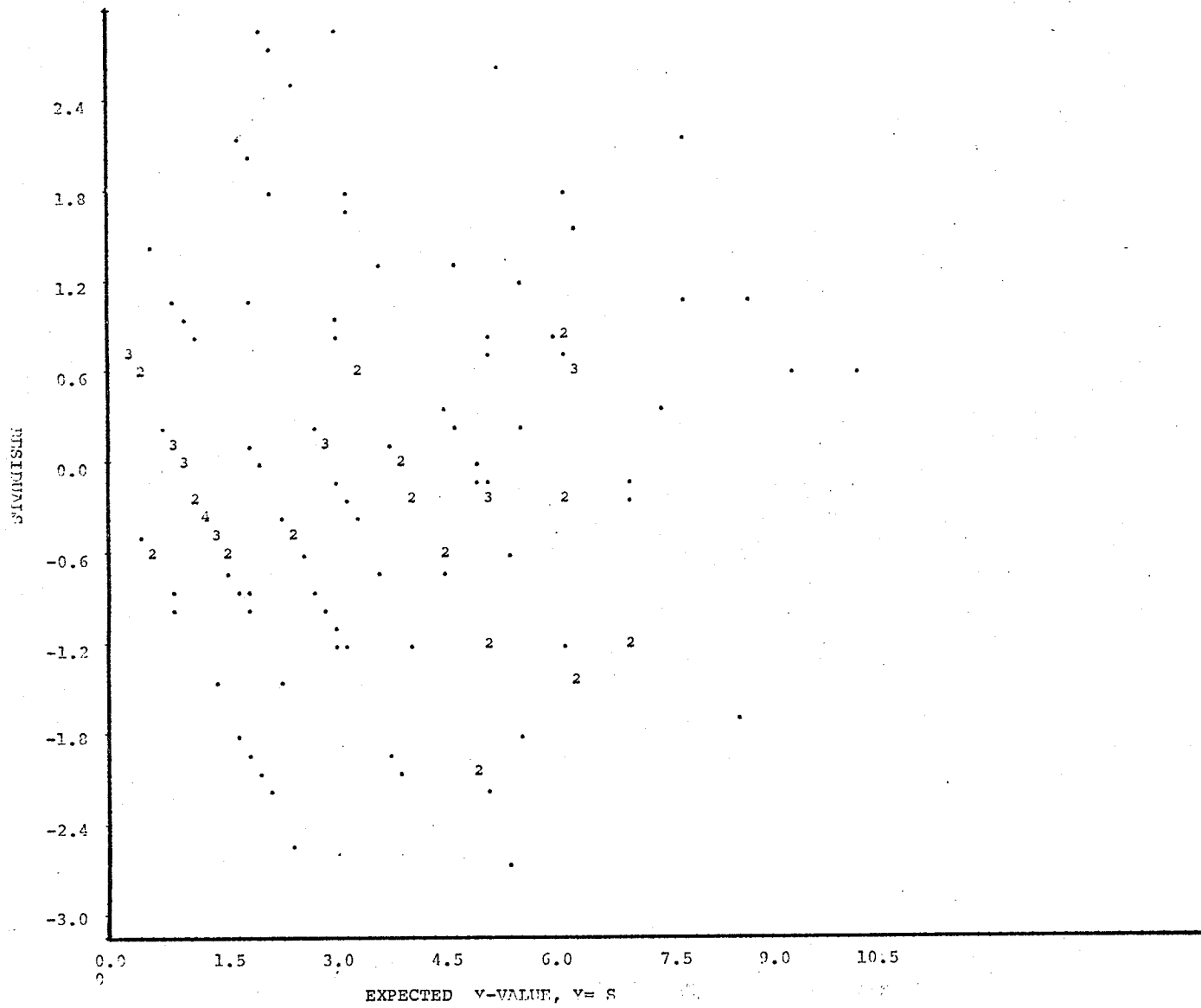
Figure 3. Plot of residuals versus expected Y-values, for Y = $\overline{)S+1}$.

RESIDUALS

0.8

0.6

0.4

0.2

0.0

-0.2

-0.4

-0.6

-0.8

0.6   0.9   1.2   1.5   1.8   2.1   2.4   2.7   3.0   3.3   3.6

EXPECTED Y-VALUE, $Y=\sqrt{S+1}$
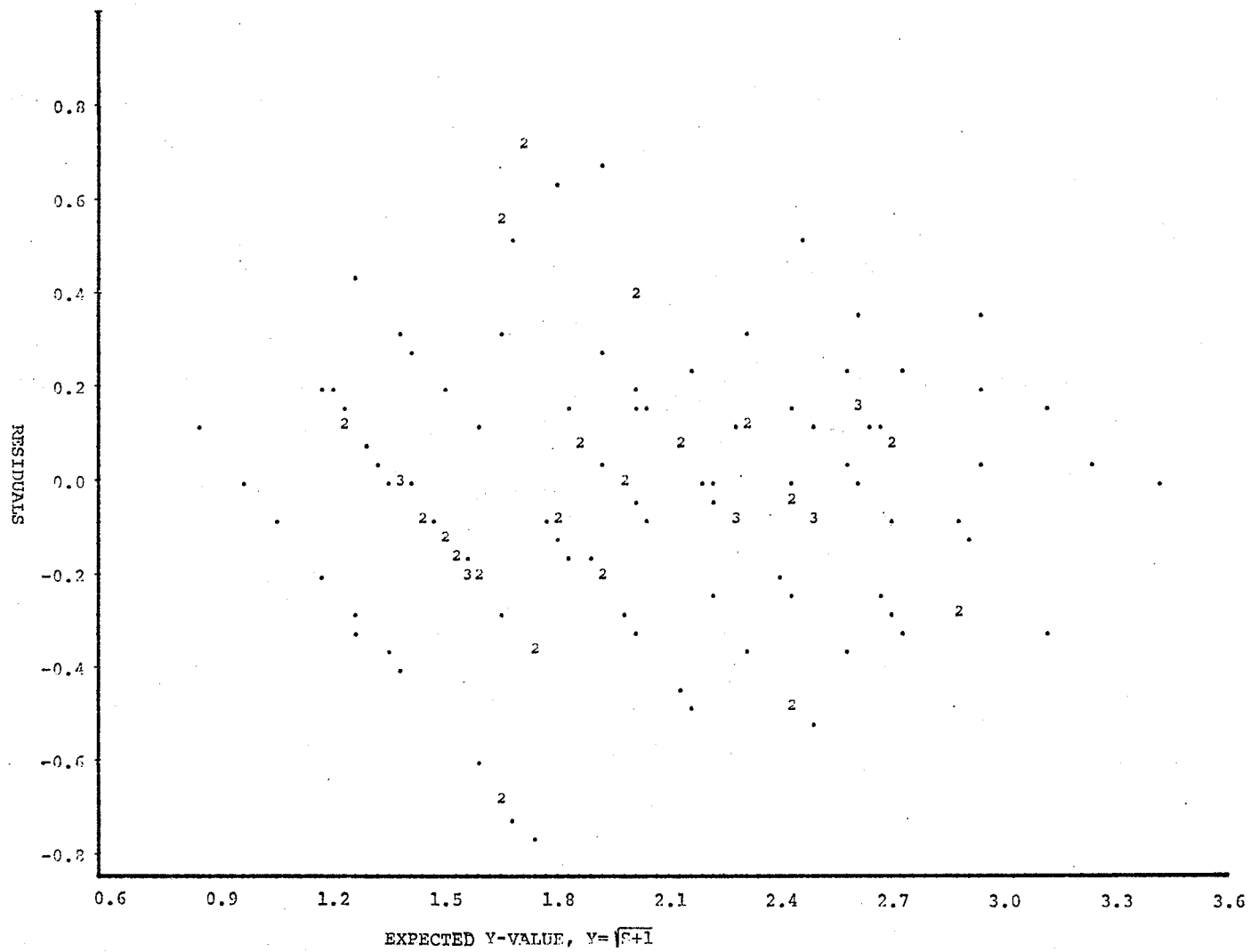
RESULTS

1. <u>Distribution of species among habitats</u>: Twenty-seven species were observed in the sampling area (Table 3). The estimated distribution of probabilities of observing each of the thirteen common species in each of the six habitat types (as defined in Table 1) is given in Table 4.

The results of the grouping procedure are given in Fig. 2. The horizontal axis is Q, and so is a measure of the heterogeneity of the groups formed. It appears from inspection of Fig. 2 that three groups of species were formed at a fairly low level of fusion, while subsequent combinations of these three were less similar. The three groups were (approximately in order of decreasing current speed preference)

1) <u>Rhinichthys</u> <u>cataractae</u> (longnose dace)

<u>R. atratulus</u> (blacknose dace)

<u>Percina</u> <u>maculata</u> (blackside darter)

2) <u>Notropis</u> <u>dorsalis</u> (bigmouth shiner)

<u>Etheostoma</u> <u>nigrum</u> (johnny darter)

<u>Catostomus</u> <u>commersoni</u> (white sucker)

<u>Semotilus</u> <u>margarita</u> (pearl dace)

<u>N. stramineus</u> (sand shiner)

<u>N. cornutus</u> (common shiner)

3) <u>Pimephales</u> <u>promelas</u> (fathead minnow)

<u>S. atromaculatus</u> (creek chub)

Table 3. Checklist of Species

	Pembina	Cypress	Oak	Plum	Neepawa	Garland
Salvelinus fontinalis*					x	
Chrosomus neogaeus					x	x
Cyprinus carpio*	x	x				
Hybognathus hankinsoni*	x					
Notropis cornutus	x	x	x			
N. dorsalis	x	x	x			
N. heterodon*			x			
N. heterolepis*			x			
N. hudsonius*			x			
N. stramineus	x	x	x			
Pimephales promelas	x	x	x	x	x	x
Rhinichthys atratulus	x	x	x	x	x	x
R. cataractae		x	x	x		
Semotilus atromaculatus	x	x	x		x	
S. margarita			x		x	x
Carpiodes cyprinus*		x				
Catostomus commersoni	x	x	x	x	x	x
Percopsis omiscomaycus*	x		x			
Esox lucius*	x		x			
Lota lota*			x			
Culaea inconstans	x		x	x	x	x
Etheostoma exile*				x		x
E. nigrum	x	x	x	x	x	x
Perca flavescens*	x		x			
Percina maculata	x	x	x			
Stizostedion vitreum*			x			
Ambloplites rupestris*			x			

* An asterisk following a species' name indicates that the species
was uncommon.

Table 4. Probabilities of Observing Each Species in Each Habitat
          Type (All Individuals)

Species	Habitat Type					
	(1)	(2)	(3)	(4)	(5)	(6)
Rhinichthys cataractae	0.61	.06	.00	.30	.03	.00
Rhinichthys atratulus	.11	.06	.42	.05	.29	.07
Percina maculata	.11	.03	.32	.25	.16	.13
Notropis dorsalis	.03	.06	.00	.59	.15	.17
Etheostoma nigrum	.06	.05	.05	.49	.13	.22
Catostomus commersoni	.01	.03	.01	.59	.21	.15
Semotilus margarita	.00	.01	.23	.40	.24	.12
Notropis stramineus	.00	.03	.08	.46	.28	.15
Notropis cornutus	.00	.02	.15	.44	.14	.25
Pimephales promelas	.00	.00	.01	.03	.88	.08
Semotilus atromaculatus	.02	.03	.03	.11	.52	.29
Culaea inconstans	.03	.02	.02	.18	.28	.47
Chrosomus neogaeus	.00	.00	.02	.17	.13	.68

Figure 4.   Dendrogram showing pattern of fusion of species into
            groups, for both size classes. Q is the within-group
            dispersion.
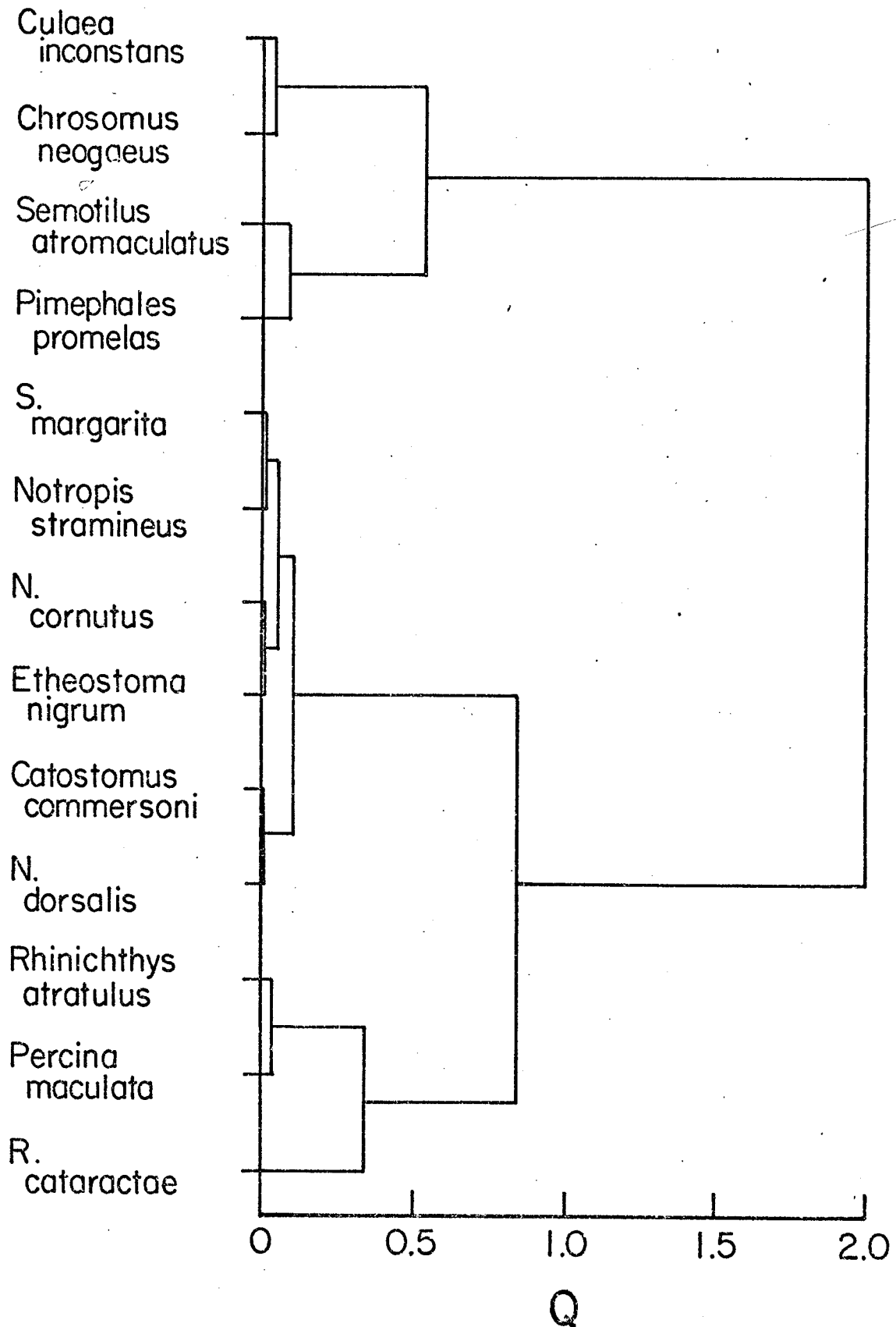
<u>Culaea</u> <u>inconstans</u> (brook stickleback)

<u>Chrosomus</u> <u>neogaeus</u> (finescale dace)

The distribution of probabilities for the larger size class alone is given in Table 5. The results of this analysis are given in Fig. 5. The groups formed were the same as in the previous case, except that the pearl dace changed its group membership from group 2 to group 3.

2. <u>Species Diversity</u>: The results of the analysis are given in Table 6. The four treatment effects, with their two-way interactions, accounted for nearly 80 percent of the variability in Y (not including "replications of sites", which is basically an error term).

DISCUSSION

1. <u>Distribution of species among habitat types</u>: By inspection of Fig. 5, it appears that three groups of species can be defined on the basis of observed distribution over habitats of the larger size class.

<u>Group 1</u>: Longnose dace, blacknose dace, blackside darter. The latter two species were most commonly found in moderate channels (type 3) and to a lesser extent in riffles and pools. Larger blackside darters seemed to be associated more with large rock substrates than with any particular current speed, at least in areas with current speed less than about 0.5 m/sec. This apparent affinity has been noted previously for <u>Percina shumardi</u>, the river darter (Trautman 1957), but Karr (1963) and Scott and Crossman (1973) have said that the blackside darter usually is found in deep pools.

Adult longnose dace were found almost entirely in riffles. Their high probability of occurrence in this habitat type separated them somewhat from the other two species in this group. This difference corresponds to the description given by Gibbons and Gee (1972) of separation between longnose and blacknose dace.

<u>Group 2</u>: Bigmouth shiner, johnny darter, white sucker, sand shiner, common shiner. The species in this group all had a high probability of occurrence in pools, and in particular in those pools located below riffles. Allocation among other habitats was somewhat variable: johnny darters were frequently found in riffles, but less

Table 5. Probabilities of Observing Each Species in Each Habitat
Type (Larger Size Class)

Species	Habitat Type					
	(1)	(2)	(3)	(4)	(5)	(6)
Rhinichthys cataractae	0.80	.00	.00	.18	.02	.00
Percina maculata	.26	.00	.35	.13	.26	.00
Rhinichthys atratulus	.18	.04	.35	.08	.25	.10
Etheostoma nigrum	.15	.08	.09	.36	.17	.15
Notropis dorsalis	.01	.11	.00	.59	.21	.08
Catostomus commersoni	.01	.03	.01	.59	.21	.15
Notropis stramineus	.00	.03	.08	.46	.28	.15
Notropis cornutus	.01	.03	.04	.46	.18	.28
Pimephales promelas	.00	.00	.01	.03	.88	.08
Semotilus margarita	.01	.00	.14	.23	.38	.24
Culaea inconstans	.08	.06	.04	.09	.44	.29
Semotilus atromaculatus	.03	.03	.04	.12	.36	.42
Chrosomus neogaeus	.00	.00	.00	.19	.10	.71

Figure 5. Dendrogram showing pattern of fusion of species into
groups, for the larger size class. Q is the within-
group dispersion.
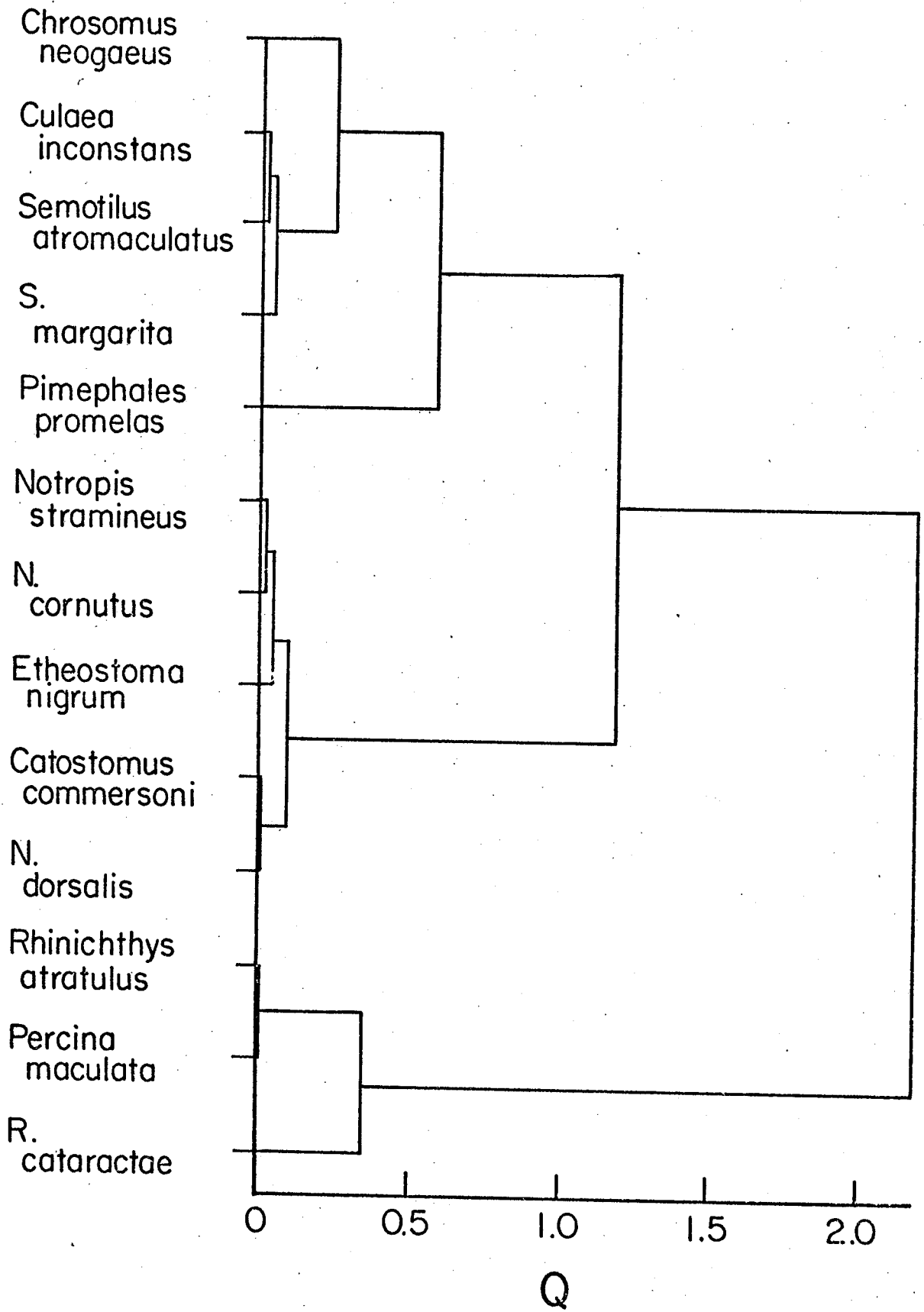
Table 6. Analysis of Variance of $Y = \overline{)S+1}$, S = Number of Species.

Source	d.f.	S.S.	M.S.(M.S.*)	F
Due to all factors	51	40.7204	--	--
Gradient (after fitting) Remaining factors	1		0.078(0.195)	0.40
Streams (")	3		0.067(.192)	0.40
Habitats (")	2		0.090(.192)	0.47
Gradients x streams (")	3		0.294(.257)	1.15
Gradients x habitats (")	2		0.399(.145)	2.74
Streams x habitats (")	6		0.136(.160)	0.85
Season (")	3		0.016	0.17
Gradients x season (")	3		0.122	1.34
Streams x season (")	9		0.144	1.58
Habitats x season (")	6		0.110	1.24
Error	73	6.6408	0.091	
Total	124	47.3612	$100R^2=86\%$	

M.S.* is the "replications of sites" error term, for factors not involving season.

often in type 2 or 3 habitats. The remaining species in the group were uniformly low in their occurrence in the first three habitat types, with the exception of the bigmouth shiner, which showed the highest probability, of any of the thirteen species, of occurrence in shallow, gravel channels with moderate to high current speed. Small groups of bigmouth shiners were often observed in this type of habitat, maintaining position near the substrate; occasionally one individual would rise up to take a drifting insect or piece of algae, and then return to its former position. Hubbs (1941) gave a similar description of habitat preference of this species. The remaining three species in this group were found mainly in areas with little current. Sand shiners, common shiners and white suckers were encountered in habitat types 4, 5 and 6 combined with probabilities of 0.89, 0.92 and 0.95, respectively.

Group 3: Pearl dace, fathead minnow, brook stickleback, creek chub, finescale dace. This group of species was separated from the previous two by its greater weight on type 5 and 6 habitats. This group was also more variable than the other two, due to the high probability for fathead minnows in type 5 habitats and the high probability of observing finescale dace in type 6 habitats. The affinity of fathead minnows for type 5 pools was largely due to the aggregated nature of the distribution of this species among stations: a single station in Neepawa Creek contained 471 fathead minnows, and the inclusion of this value in calculating the p_{ij} greatly diminished the relative contribution (to the calculated probabilities) of a

large number of other stations in which this species was less
abundant. Finescale dace were at no time very abundant, but were
quite consistent in occurring in deep, slow channels. Also, in the
streams in which members of this species were present, they were
mainly confined to the low gradient sections. Creek chub and
sticklebacks were similar in their allocation to habitats, the latter
species showing a somewhat higher weight on type 5 pools. Sticklebacks,
however, were found in greatest abundance in low gradient sections,
while creek chub were found in similar habitats but within the high
gradient sections. Pearl dace were somewhat more variable in habitat
selection than the other species in this group, occurring with fairly
high probability in habitats 3 and 4 (moderate channels and below
riffles, respectively) as well as types 5 and 6. Scott and Crossman
(1973) said that pearl dace, finescale dace, brook sticklebacks and
fathead minnows are often found in the same streams; it appears that
this association may also occur at a level of habitats within streams
as well as at a level of streams.

By inspection of the p_{ij} in Tables 4 and 5, it can be seen
that removal of the smaller size class changed the apparent habitat
preference of some species more than others. The greatest change
was observed for those species occurring in higher current speeds as
adults, which is an intuitively reasonable result, since smaller
individuals probably have a similar lack of ability to resist strong
current, irrespective of species. Longnose dace, blacknose dace
and blackside darters seemed to change habitat preference by about

the same amount, and in the same direction; thus in both cases these three species formed a group at about the same level of heterogeneity. In the first case, however, (smaller size class included) this group of species was closer to the pool group and joined with the pool species at a lower level of fusion. Changes were smaller in the pool and deep channel groups: the only change in group membership was the pearl dace, which moved from the pool group, when all individuals were considered, to the deep channel group when only the larger size class was considered. Even this change was small, however, as this species when included in the channel group had the highest probability, for that group, of occurrence in pools.

2. Species Diversity: A discussion of the factors included in the analysis of variance follows:

Gradient: The number of species in low gradient stream sections was considerably less (Table 9) than in high gradient sections, and only one species, the brook stickleback, was very abundant in low gradient areas. Possible explanations for this difference include differences in temperature regime between the low and high gradient areas, mechanical barriers to colonization or differences in food availability. Riffles are considered to be the most productive areas in streams (Needham 1934) in terms of the invertebrates upon which many stream fishes feed, and few such environments occurred in low gradient sections. Riffles are also the major source of drifting invertebrates (Waters 1972) and Mason and Chapman (1965) found that the carrying capacity of streams for fish

was related to the amount of drift in the streams. Alternatively, food may be scarce in low gradient stream sections; Hynes (1970) suggested that the substrate of a low gradient stream presents an environment for invertebrates which is similar to a lake bottom, except that the stream bed is unstable due to variations in discharge. Thus, while the silt substrate of a low gradient stream may at times be a suitable environment for lacustrine species of invertebrates, the stream bed is generally not stable long enough for suitable species to colonize and increase in abundance.

Burton and Odum (1945) suggested that temperature limited the distribution of stream fish species in Virginia, but maximum temperatures recorded in this study did not differ in high and low gradients, so it appears unlikely that temperature differences were exerting an effect on diversity of fish species.

Possible mechanical barriers in the form of beaver or man-made dams were present in two streams, but similar diversity differences were observed in the three streams without dams. It also seems unlikely that the low gradient sections were effectively isolated by distance, since many of the fishes which did not appear in the low gradient areas have been noted for their colonization ability, for example the creek chub (Larimore, Childers and Heckrotte 1959).

Habitats: The effect of habitats appeared to be due mainly to a difference between pools and shallower areas (Table 7), pools containing a higher average number of species. This increase in

diversity of species across the riffle-pool spectrum has been
described in several other studies, such as those of Gard and
Flittner (1974), Minckley (1963) and Sheldon (1968). Since depth
and current speed are negatively correlated in streams, it is difficult
to determine experimentally whether the difference between pools and
riffles was due to differences in depth, current speed or both,
since these two factors could not be varied independently. The first
two studies above suggested that the fishes responded to current speed,
while Sheldon suggested that increased depth may result in increased
habitat diversity, in that the fishes may occupy different depth
niches with respect to feeding behaviour. Exactly this result has
been described in a tropical stream by Zaret and Rand (1971). However,
while it may be true that pools are structurally more diverse than
riffles, this habitat diversity is probably due more to current
speed differences than to depth differences. In this study, some fast,
shallow channels which contained no species early in the summer were
colonized later on when discharge decreased, with final species
number approaching that of deeper pools.

In a riffle, if a fish is not to be swept away, it must
either resist the current by actively swimming or avoid it by staying
near the substrate. Of these two strategies, the second would seem
to be superior, since the former would require a large energy
expenditure just to maintain position, and would probably require a
higher intake of food to meet this energy demand. Also, there is an
upper limit to the current speed against which any fish can hold

position; it is thought (Alexander 1967; Bainbridge 1960) that no fish can swim indefinitely at more than three to five lengths per second. Thus, any fish species which inhabits fast riffles is probably obligated to make use of the substrate to avoid the current. From a point of view of food supply, the same restrictions apply to invertebrates in running water as to fishes, and the riffle dwelling invertebrates are therefore necessarily benthic.

In pools, however, with the restriction due to current speed removed, the fish species are free to occupy different spatial niches. In this study, white suckers and some darter species were primarily benthic in pools, while other pool dwelling species appeared to occupy the water column to varying extents. Specialization for particular feeding depths within the water column does not appear to be as important in temperate waters as it does in some tropical streams, such as the one studied by Zaret and Rand (1971), but at least some temperate stream species are known to be specialized in this way, such as the surface feeding redbelly dace, _Clinostomus elongatus_.

Seasonal changes: The main effect due to season appeared small (Table 6) relative to the interactions between season and some other factors. This is a reflection of the fact that the increase in average number of species per station over the summer was small relative to seasonal changes in the effects of the other factors.

Interactions:

Season x gradients: High gradient areas increased in average species number in the latter part of the summer, while low gradient

areas showed little increase (Table 8). This effect appeared to be largely due to colonization of the lower section of Oak Creek by species which occur more commonly in larger bodies of water (e.g. the yellow perch, _Perca flavescens_; rock bass, _Ambloplites rupestris_; burbot, _Lota lota_). In Oak Creek, stations were located near the outlet, whereas in the other streams the lowest stations were farther upstream.

Gradients x streams: This interaction (Table 9) again appeared to be due to the Oak Creek high gradient section having more species present than predicted on a basis of the difference due to streams. This interaction and the previous one appeared to be both due to a large number of species in the lower part of this stream.

Habitats x season: The difference between habitats decreased throughout the summer (Table 7), due to colonization of shallow stations as the current speed decreased. This effect was mentioned in connection with the discussion of habitats and would seem to suggest that current speed limited species number more than depth did.

Streams x habitats: The streams x habitats interaction was the least significant of the interaction terms when tested as shown in Table 6. Since the streams in the study area did not contain the same species, it was anticipated that the relative species numbers in the three habitats would differ more than they did from stream to stream (Table 10). This suggests that the increase in number of species across the riffle-pool spectrum was reasonably consistent in these streams. The large number of references that have been made

to this effect in other studies suggest that this is a general result.

Gradients x habitats: High gradient sections (Table 11) showed a greater difference between riffles and pools than low gradient sections did.

Streams x time: This interaction appeared to be due mainly to the lack of increase in number of species, with time, in Plum Creek (Table 12). Plum Creek partially dried up by July, 1973.

Summarizing the analysis of variance results, it would appear that variability in number of fish species was largely explained by gradient and by habitat type, after accounting for differences in the number of species in each stream. The effects of gradient and habitat changed over time and in different streams, however, and as a result the interactions appeared to be larger than the main effects. Seasonal changes were evident, first by colonization of Oak Creek in late summer, and second by colonization of shallow areas by species already present when stream discharge decreased.

The observed effects of environmental factors agree in general with those described in other studies. The greatest difference in results is the apparent lack of longitudinal succession (Shelford 1911) observed in this study. In many streams studies elsewhere (Burton and Odum 1945; Minckley 1963; Sheldon 1968; Whiteside and McNatt 1972), a pattern has been observed whereby species are progressively added as distance from the headwaters increases. Succession of this type has usually been attributed to increased

Table 7. Mean Y-value Under Each Combination
of Habitats and Seasons

	Channel	Riffle	Pool
May	1.642	1.372	2.076
June	1.836	1.483	2.265
July	1.901	1.815	2.371
August	2.124	2.248	2.507

Table 8. Mean Y-value Under Each
Combination of Gradient
and Season

	LG	HG
May	1.455	1.961
June	1.607	2.216
July	1.650	2.401
August	1.801	2.669

Table 9. Mean Y-value Under Each Combination of Stream
          and Gradient

	Garland	Neepawa	Oak	Pembina	Plum
LG	1.389	1.766	1.569	2.208	1.441
HG	2.108	2.378	2.688	2.608	2.014

Table 10. Mean Y-value Under Each Combination of Habitat and Stream

	Channel	Riffle	Pool
Garland	1.693	1.329	2.013
Neepawa	1.981	1.651	2.341
Plum	1.467	1.414	1.842
Oak	2.018	2.390	2.471

Table 11. Mean Y-value Under Each Combination
           of Habitat and Gradient

	Channel	Riffle	Pool
LG	1.487	1.293	1.734
HG	2.197	1.961	2.663

Table 12. Mean Y-value Under Each Combination of
Stream and Season.

	Garland	Neepawa	Plum	Oak
May	1.488	1.771	1.685	1.843
June	1.571	2.002	1.494	2.084
July	1.558	1.816	1.620	2.506
August	2.096	2.346	1.620	2.536

habitat diversity in downstream areas as a result of increasing
stream size, and in fact Thompson and Hunt (1930) used a measure of
stream size to predict the number of fish species present. An
increase in habitat diversity may occur, but it is also possible that
the increase in species diversity described is at least partly due to
increased sampling area. Emlen (1973) has shown that even in homogeneous
habitats, observed species number will increase with sampling area if
the species are distributed at random. At any rate, the streams in
the study did not increase noticeably in size along their lengths,
and they were also relatively short, which may explain the lack of
longitudinal succession.

Biotic Factors: Pike (Esox lucius) when present had a large
effect in reducing species diversity of the remaining species. In
early July, 1974, large numbers of pike fry were observed drifting
down Oak Creek towards the Souris River. This was not observed the
previous year and the large numbers may have been due to extensive
marsh flooding in 1974. The drifting pike were seen feeding on common
shiners and creek chub, in some cases nearly as large as the pike
themselves. When the stream was next sampled, on August 9, a few
pike remained in the pools, while both diversity and abundances of
the remaining species were greatly reduced. Average numbers of each
species, before and after the pike were present, and in August, 1973,
are given in Table 13 for riffles, channels and pools separately. It
can be seen that some species were reduced in abundance more than
others: the greatest changes in abundance occurred for creek chub,

common shiners and white suckers, while the riffle dwelling species, particularly longnose dace, seemed to be unaffected. Bigmouth shiners also remained abundant, but mainly in fast, shallow channels.

The August, 1974, collections show even greater differences when compared with the August, 1973, collections. On the basis of 1973 collections, some colonization was expected in Oak Creek in August. Also, (the habitat x season interaction) riffles in 1973 increased in species number with reduced current speed. This colonization effect was not observed in 1974 when populations in pools adjacent to the riffles were low, and this suggests that a high species number in riffles may depend on colonization from adjacent pools.

A similar effect was observed in the Pembina River. In those deep pools where pike were found, both diversity and density of other species were low. It would seem that the prey species were actively avoiding areas where pike were present, since in other parts of the same stream, stations from which all fish were removed in the course of normal sampling were recolonized in a matter of minutes when the barrier nets were removed.

The observed effect of a predator on community structure is in contrast to the effect observed by Paine (1966). Paine removed predators (the starfish, Pisaster sp.) from sections of intertidal beach, and found that in its absence some species of barnacles and later, bivalves, were able to increase in abundance and crowd out remaining species, with the result that the number of species present

46

Table 13. Collections Made in Oak Creek in Presence and Absence of Pike

| | Number of each species (average of two stations) | | | | | | | | |
| | July, 1974 | | | August, 1974 | | | August, 1973 | | |
	R*	C	P	R	C	P	R	C	P
Rhinichithys cataractae	52	--	--	56.5	--	--	45	48.5	--
Rhinichthys atratulus	14	--	--	5	--	--	1	6	2
Notropis cornutus	--	8	58	--	--	0.5	7	13.5	10
Notropis dorsalis	--	0.5	2	--	8.5	1	2	11	18
Notropis heterodon	--	--	--	--	--	--	--	--	1
Notropis heterolepis	--	--	--	--	--	--	--	--	1
Notropis hudsonius	--	--	--	--	--	--	--	0.5	--
Pimephales promelas	--	--	--	--	--	--	--	--	1
Semotilus atromaculatus	--	4	15.5	--	--	--	5	16	24
Semotilus margarita	--	--	0.5	--	--	--	--	0.5	--
Catostomus commersoni	1.5	3	46.5	--	--	--	2	3	6
Culaea inconstans	--	--	--	--	--	--	--	--	2
Etheostoma nigrum	3.5	3.5	2	1.5	1	0.5	--	8.5	4
Percina maculata	1.5	0.5	2.5	0.5	--	0.5	1	1.5	1
Ambloplites rupestris	--	--	1	--	--	--	--	--	--

dropped from fifteen to eight. The reason for the difference in effects of the two predators is probably that the intertidal species colonized wherever space was available, independently of presence or absence of the predator, while the fish species considered here seemed to actively avoid areas where pike were present.

Few similar predation effects have been mentioned in other stream fish studies (although see Farr 1975), but biotic factors in general seem to be fairly important, particularly in tropical streams. As stated previously, Zaret and Rand (1971) found that in a Panamanian stream, the fish species occupied different depth niches with respect to feeding behaviour, particularly in the dry season when water levels were low and available space was reduced. Hutchinson (1939) found a negative association among Himalayan loaches, in which the species tended to replace one another along lengths of stream.

A similar situation was described by Lachner, Westlake and Handwerk (1950) for some darter species, but in temperate areas the evidence for competition seems to be largely restricted to single species pairs, or to one introduced species versus a set of resident species. Kawanabe (1969) found that fishes resident in certain Japanese streams were displaced, both in food habits and in distribution, when the anadromous ayu, _Plecoglossus altivelis_, was present. Trautman (1957) stated that the silverjaw minnow, _Ericymba buccata_, displaced the ecologically similar bigmouth shiner when the first species invaded streams occupied by the second. Deacon and Bradley (1972) found that stream species native to Nevada were displaced after introduction of the cyprinodont, _Poecilia mexicana._

SUMMARY AND CONCLUSIONS

In summary, the within-stream distribution of species was described and species diversity was related to predation and environmental parameters. Species diversity was low in low gradient areas and increased across the riffle-pool spectrum. Diversity was reduced when a predator was present. Seasonal changes consisted largely of first, colonization of the lower parts of two streams and second, movement of species into shallow areas as stream discharge decreased.

The conclusion drawn from this study is that gradient, current speed and predation, in about that order of importance, largely determined species diversity in the streams considered.

REFERENCES

Alexander, R. McN. 1967. Functional design in fishes. Hutchinson
     and Co. Ltd., London. 160 p.

Armitage, P. A. 1971. Statistical methods in medical research.
     Wiley, New York, N. Y. 504 p.

Bainbridge, R. 1960. Speed and stamina in three fish. J. Exper.
     Biology 37: 129-153.

Brillouin, L. 1956. Science and information theory. Academic
     Press, Inc., New York, N. Y. 320 p.

Burton, G. W. and E. P. Odum. 1945. The distribution of stream fish
     in the vicinity of Mountain Lake, Virginia. Ecology 26: 182-194.

Daniel, C. and F. S. Wood. 1971. Fitting equations to data. Wiley,
     Inc., N. Y. 342 p.

Deacon, J. E. and W. G. Bradley. 1972. Ecological distribution of
     fishes in the Moapa River in Clark County, Nevada. Trans. Amer.
     Fish. Soc. 101: 408-419.

Draper, N. R. and H. Smith. 1966. Applied regression analysis.
     Wiley, New York.

Emlen, J. M. 1973. Ecology: an evolutionary approach. Addison-
     Wesley Publ. Co., Reading, Mass. 493 p.

Farr, J. A. 1975. The role of predation in the evolution of natural
     populations of the guppy *Poecilia reticulata* (Pisces: Poeciliidae).
     Evolution 29: 151-158.

Gard, R. and G. A. Flittner. 1974. Distribution and abundance of
     fish in Sagehen Creek, California. J. Wildl. Man. 38: 347-358.

Gibbons, J. R. H. and J. H. Gee. 1972. Ecological segregation between
     longnose and blacknose dace in the Mink River, Manitoba. J. Fish
     Res. Bd. Canada 29: 1245-1252.

Good, I. J. 1953. The population frequencies of species and the
     estimation of population parameters. Biometrika 40: 237-264.

Hubbs, C. L. 1941. The relation of hydrological conditions to
     speciation in fishes. p. 182-195. In a symposium on hydrobiology.
     Univ. Wisconsin Press, Madison, Wisc.

Hurlbert, S. H. 1971. The nonconcept of species diversity: a
     critique and alternative parameters. Ecology 52: 577-586.

Hutchinson, G. E. 1939. Ecological observations on the fishes of
Kashmir and Indian Tibet. Ecol. Monographs 9: 145-182.

_____. 1959. Homage to Santa Rosalia, or Why are there
so many kinds of animals? Amer. Natur. 93: 145-159.

Hynes, H. B. N. 1970. The ecology of running waters. Univ. of
Toronto Press. 555 p.

Karr, J. R. 1963. Age, growth and food habits of johnny, slenderhead
and blacksided darters of Boone County, Iowa. Proc. Iowa Acad.
Sci. 70: 228-236.

Kawanabe, H. 1969. The significance of social structure in production
of the "Ayu", Plecoglossus altivelis, p. 1-10. In T. G.
Northcote (ed.) Salmon and trout in streams. H. R. MacMillan
lectures in fisheries. Univ. of British Columbia, Vancouver, B. C.

Kullback, S. 1959. Information, theory and statistics. Dover Publ.,
Inc. New York, N. Y. 399 p.

Lackner, E. A., E. F. Westlake and P. S. Handwerk. 1950. Studies on
the percid fishes from western Pennsylvania. Amer. Midland
Natur. 43: 92-111.

Larimore, R. W., W. F. Childers and C. Heckrotte. 1959. Destruction
and re-establishment of stream fish and invertebrates affected
by drought. Trans. Amer. Fish. Soc. 88- 261-285.

Mason, J. C. and D. W. Chapman. 1965. Significance of early emergence,
environmental rearing capacity and behavioral ecology of juvenile
coho salmon in stream channels. J. Fish. Res. Bd. Canada 22:
173-190.

McIntosh, R. P. 1967. An index of diversity and the relation of
certain concepts to diversity. Ecology 48: 392-404.

Minckley, W. L. 1963. The ecology of a spring stream: Doe Run,
Meade County, Kentucky. Wildl. Monogr. 11: 124 p.

Needham, P. R. 1934. Quantitative studies of stream bottom foods.
Trans. Amer. Fish. Soc. 64: 238-247.

Orloci, L. 1967. An agglomerative method for classification of plant
communities. J. Ecology 55: 193-205.

Paine, R. T. 1966. Food web complexity and species diversity.
Amer. Natur. 100: 65-75.

Pielou, E. C. 1966. The measurement of diversity in different
    types of biological collections. J. Theor. Biology 13: 131-144.

Scott, W. B. and E. J. Crossman. 1973. Freshwater fishes of Canada.
    Fish. Res. Bd. Canada Bull. 184: 966 p.

Shannon, C. E. 1948. A mathematical theory of communication. Bell
    System Tech. J. 27: 379-423; 623-656.

Sheldon, A. L. 1968. Species diversity and longitudinal succession
    in stream fishes. Ecology 49: 193-197.

Shelford, W. E. 1911. Ecological succession. I. Stream fishes and
    the method of physiographic analysis. Biol. Bull. 21: 9-35.

Snedecor, G. W. and W. G. Cochran. 1967. Statistical methods.
    Iowa State Univ. Press, Ames, Iowa. 593 p.

Steel, R. G. D. and J. H. Torrie. 1960. Principles and Procedures
    of Statistics. McGraw-Hill, Inc. New York, N. Y. 481 p.

Thompson, D. H. and F. D. Hunt. 1930. The fishes of Champaign
    County. Ill. Natur. Hist. Surv. Bull. 19: 5-101.

Trautman, M. B. 1957. The fishes of Ohio. Ohio State Univ. Press,
    Columbus, Ohio. 683 p.

Waters, T. F. 1972. The drift of stream insects. Ann. Rev. Entomol.
    17: 253-272.

Whiteside, B. G. and R. M. McNatt. 1972. Fish species diversity in
    relation to stream order and physiochemical conditions in the
    Plum Creek drainage basin. Amer. Midland Natur. 88: 90-101.

Zaret, T. M. and A. S. Rand. 1971. Competition in tropical stream
    fishes: support for the competitive exclusion principle.
    Ecology 52: 336-342.

APPENDICES

1.   INDICES OF SPECIES DIVERSITY

Biological communities differ both in the number of species they contain and in the relative abundance of the component species. In order to compare the structure of different communities or to relate community structure to characteristics of the environment, it would be useful to have a descriptive statistic which summarized the information contained in these two parameters.

Several indices have been proposed as measures of species diversity (for example, Simpson 1949; Good 1953; McIntosh 1967).

The most widely used index at present is probably the Shannon-Wiener function.

$$H(s) = -C\sum_i p_i \log(p_i) \qquad (1) \quad [\text{Shannon 1948; Wiener (in}$$

$$\text{Kullback 1959)}]$$

where S = the number of species present,

p_i = probability of observing the (i)th species,

$i = 1, 2, \ldots, S$.

Because of the widespread application of this diversity index, it will be used as a basis for the following general discussion of diversity indices.

In terms of information theory, the information content of a sample of organisms depends on the uncertainty associated with the specific identity of a selected individual. If all individuals belong to different species, then the uncertainty is large and hence the information content. If all individuals belong to the

same species then uncertainty, and hence information, is zero. A

more precise definition was summarized from Kullback (1959):

Consider a sample value, X, of a variable X. If there exist

two hypotheses of interest, say

H1: X has probability function f(x)

H2: X has probability function g(x)

and it is desired to decide which of H1, H2 is more likely, then

by Bayes' rule,

$$Pr(H1/x) = [Pr(H1)f(x)] / [Pr(H1)f(x) + Pr(H2)g(x)]$$

Rearranging,

$$F(x)/g(x) = [Pr(H1/x) / Pr(H2/x)] / [Pr(H1) / Pr(H2)] \quad (2)$$

where $f(x)/g(x)$ is just the ratio of the two alternative probability

functions, or likelihood ratio, $L(x)$ used in ordinary hypothesis

testing [H1 is rejected if (L(x) under H1 / (L(x) under H2) is less

than 0.05, usually]. Taking logs of both sides, (2) becomes

$$log[f(x) / g(x)] = log(Pr(H1/x) / Pr(H2/x) - log[Pr(H1) / Pr(H2)]$$
$$(3)$$

= the difference between the logs of the probability ratios,

before and after observing x.

The log likelihood ratio (3) is defined as the information, 1, in the

observation x, for choosing H1 over H2. Taking logs is just for

convenience; it makes the information function additive without changing

any of the properties of the likelihood ratio, since $z = log(y)$ defines

a one to one transformation from y to z.

As a more specific example, suppose that H2 represents a set

of hypotheses, one of which must be true and that H1 is a single

member of the set H2; then $Pr(H2) = 1$, and $Pr(H2/x) = 1$. If, from the observation, it can be established with certainty that H1 is true (for example H1: this individual belongs to species A), then $Pr(H1/x) = 1$. Then 1 becomes

$$1 = -\log[Pr(H1)]$$

Further, if it is known that one of Hi: an individual belongs to species (i), i = 1, 2, ..., S is true, then the information in any individual about Hi is $-\log[Pr(Hi)$; the mean, or expected information in any individual about the hypotheses is

$$H(s) = -\sum_i Pr(Hi)\log[Pr(Hi)]$$

the Shannon-Wiener function.

Brillouin (1956) used this definition of information, but he also attempted to give a more intuitive derivation by comparing the information content of a sample to the logarithm of the multinomial coefficient: for a sample of size N objects, of which N1 are of one kind, ... Ns are both of the (s)th kind,

$$H(b) = \log(N! \ / \ N1! \ N2! \ ... \ Ns!).$$

It seems worth mentioning that he did this for heuristic purposes only. H(b) and H(s) are approximately the same for large N, Ni, but are not close for small samples. The true information content is always H(s), and H(b) is an approximation to H(s), but the converse is not true.

Pielou (1966) proposed that Brillouin's index be used as the measure of information in a finite sample (that is, when considering the diversity of the sample itself), and that H(s) be used when estimating a population diversity value from the observed sample

diversity. Her reasoning for this was that H(s) only has meaning
for infinite collections, or estimates thereof, since only in the
limit as N approaches infinity does H(b) equal H(s). This is in
contradiction to Brillouin's original statement that H(b) is an
approximation to H(s).

In point of fact, this distinction is not important, since
neither index means a great deal in biological terms. The behaviour
of any diversity index depends on how the relationship between diversity
and relative abundance is defined, and this definition is necessarily
arbitrary, since in general there is no simple biological mechanism
explaining the relative abundance of species in collections (Hurlbert
1971). Thus, it is probably more important, for purposes of comparison,
that different studies use the same diversity index, than that a
particular index is chosen as "best." Using the species number S,
however, does have the advantage that its biological meaning is
relatively clear: it is a measure of the number of distinct niches
which overlap in the sample unit, in the sense of Hutchinson (1959).

2. THE ANALYSIS OF VARIANCE MODEL

The number of observations made within each treatment combination is given in Table 14. Because this number was not constant, the formula normally used to calculate the ANOVA sums of squares was not applicable. It was necessary to use a more general regression approach, which will be illustrated by use of the following example, a completely randomized analysis of variance design with two treatments and three replications of each treatment.

The fixed-effects analysis of variance model for this design is:

$$Y_{ij} = \mu + T_i + \varepsilon_{ij}, \quad i = 1, 2$$
$$j = 1, 2, 3$$

where μ is the overall mean, T_i is the effect of the ith treatment and ε_{ij} is the error associated with the jth observation under the ith treatment.

This ANOVA model can be shown to be exactly equivalent to the regression model:

$$Y = \beta_o X_o + \beta_1 X_1 + \varepsilon \qquad (\beta_o \text{ is the Y-intercept, } X_o = 1)$$

if X_1 is defined as:

$X_1 = 0$, if the observation was made under treatment 1,

$\quad = 1$, otherwise.

Although there are two treatments, only a single X-variable, X_1, is necessary to specify which treatment is applied. This is because the analysis of variance model contains more parameters than are actually needed (T_1, T_2) to measure a single difference between

two groups. (This redundancy is usually overcome in analysis of variance by applying the restriction $\Sigma T_i = 0$). In general, a factor at N levels may be defined by (N-1) X-variables.

From the regression model,

if $X = 0$, $Y = \beta_o + \varepsilon$

if $X = 1$, $Y = \beta_o + \beta_1 + \varepsilon$.

This implies that β_o represents the effect of treatment 1 and represents the difference between treatment 1 and treatment 2.

The regression calculations are as follows:

The ANOVA layout of data is

$$
\begin{array}{c|ccc}
T_1 & Y_{11} & Y_{12} & Y_{13} \\
\hline
T_2 & Y_{21} & Y_{22} & Y_{23}
\end{array}
$$

representing these data as matrices of X's and Y'x.

$$
\begin{array}{cc}
X_o & X_1
\end{array}
$$

$$
\underline{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}
\qquad
\underline{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \end{bmatrix}
$$

$$
(\underline{X}^1 \underline{X}) = \begin{bmatrix} 3 & 3 \\ 6 & 6 \end{bmatrix}
\qquad
(\underline{X}^1 \underline{X})^{-1} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 1/3 \end{bmatrix}
\qquad
(\underline{X}^1 \underline{Y}) = \begin{bmatrix} \sum\limits_{i=1}^{2} \sum\limits_{j=1}^{3} YY \\ \sum\limits_{i=1}^{2} Y_{i2} \end{bmatrix}
$$

Now, estimates of β are obtained by solving for values of __b__ which satisfy:

$\underline{Y} = \underline{Xb}$ (underlining is used to represent matrices)

Therefore

$X^1 Y = X^1 Xb$

$(X^1 X)^{-1} X^1 Y = \underline{b}$

Therefore, in this example,

$$\underline{b} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \cdot \underline{Y}$$

$$= \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 - \bar{y}_1 \end{bmatrix}$$

This result is what would be expected from the way the model was stated. The intercept is estimated by $\bar{y}_1$ and β_1 is estimated by the difference between the averages for each treatment.

The sum of squares due to regression is

$$SS_R = \underline{b}' \underline{X}' \underline{Y} - \frac{1}{6} \underline{Y}' \underline{Y} \qquad \text{(Draper and Smith, 1960)}$$

$$= \frac{1}{3} \left(\sum_i y_{11}{}^2 + \sum_i y_{12}{}^2 \right) - \frac{1}{6} \left(\sum_i \sum_j y_{ij} \right)^2$$

which is identical to the sum of squares due to treatments obtained by the usual ANOVA calculations.

The model used in this study was identical in derivation to the above example:

$$Y = \beta_o X_o$$

$+ \{\beta_1 X_1 + \dots \beta_3 X_3\} \quad \text{(four streams)}$

$+ \{\beta_4 X_4\} \quad \text{(two gradients)}$

$+ \{\beta_5 X_5 + \beta_6 X_6\} \quad \text{(three habitats)}$

$+ \{\beta_7 X_7 + \dots \beta_9 X_9\} \quad \text{(four times)}$

$+ \{\beta_{10} X_{10} + \dots \beta_{12} X_{12}\} \quad$ (three terms representing streams x gradients interaction; $X_{10} = X_1 X_4$, etc.

.

.

.

remaining interaction terms.

$$\vdots$$

$+ \{\beta_{40}X_{40} + \ldots \beta_{52}X_{52}\}$ "replications of sites" error.

There were thirteen terms representing replications of sites error. The treatment combination "Neepawa low-gradient channels" will serve as an example of how these terms were obtained. From Table 14, there were three sites representing this treatment in May, followed by three, one and two in June, July and August, respectively. Therefore, two X-variables were used to specify which of the three different sites an observation came from.

<u>Tests of Hypotheses</u>: To test the hypothesis that the effect of say, gradient was zero, i.e. H_o: $\beta_4 = 0$,

1. a model which omitted both X_4 and X_{40} - X_{52} was fitted.
2. a model which omitted only X_4 was fitted, and the additional sum of squares accounted for by this second model was used to obtain the mean square due to replications of sites error.
3. finally, the complete model was fitted, and the addition to the regression sum of squares over the second model was attributed to gradient.

The mean square due to gradient was then divided by the mean square due to replications of sites, and this ratio was assumed to have an F-distribution with one and thirteen degrees of freedom.

If the sum of squares for gradient had been obtained first, (i.e. 1, then 3, then 2), a larger F-value would have resulted, because of the lack of independence between the two terms. The more conservative

procedure was adhered to. However, by following this procedure, the replications of sites error was not constant, but changed depending on the identity of the terms present in model 1.

Terms involving the "time" factor were tested against the residual error (about the full model, after accounting for replications of sites error).

In split plot designs, the whole unit (or "sites" in this case) factors are generally estimated with less precision than the split unit ("times") factors (Steel and Torrie, 1960). In this analysis, this was an unfortunate consequence of the design, since differences among station types were of greater interest than differences due to seasonal changes.

Finally, there is some justification for a somewhat different analysis of this data. First, tests of individual factors against the reduced model are not tests of significance of effects of those factors as such, but are tests of significance of additional effects of those factors, after fitting the reduced model. Because of correlations among the factors, the additional sum of squares accounted for by any factor after fitting the reduced model is much less than the sum of squares accounted for in a model containing that factor alone. For example, a simple model containing gradient alone accounted for some 36 percent of the variability in Y.

Second, the argument can be made that main effects should have been tested against reduced models which did not contain any interaction terms for the main effects in question, since zero main

effects can have nonzero interactions only in the unlikely case in which the main effect of say, A is real, but is reversed over a factor B at two levels, such that the average effect of A is zero.

Two-way Tables: Because of the lack of orthogonality, the usual ANOVA procedure of obtaining two-way tables from the row and column totals could not be used. Therefore, each cell was expressed as an average, and then the two-way tables were calculated from these averages to eliminate the effect of different numbers of observations in each cell. This is an extension of a procedure advocated by Snedecor and Cochran (1967) (p.473), and would appear to give unbiased estimates since, for a two-way analysis of variance with factors A and B:

	B_1	B_2
A_1	$\bar{Y}_{11\cdot}$	$\bar{Y}_{12\cdot}$
A_2		

with model $Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$,

$$\frac{E(\bar{Y}_{11\cdot} + \bar{Y}_{12\cdot})}{2} = \frac{1}{2} E(\bar{Y}_{11\cdot}) + \frac{1}{2} E(\bar{Y}_{12\cdot})$$

$$= \frac{1}{2} (\mu + \alpha_1 + \beta_1) + \frac{1}{2} (\mu + \alpha_1 + \beta_2)$$

$$= \mu + \alpha_1 + \frac{(\beta_1 + \beta_2)}{2}$$

$$= \mu + \alpha_1$$

= main effect of A at level 1 since $\beta_1 + \beta_2 = 0$,

which is the desired result.

Table 14. Number of observations under each treatment combination.

Time

	LG			HG		
	C	R	P	C	R	P
Garland	1	1	1	2	2	2
Neepawa	3	1	2	1	2	3
Plum	1	1	1	1	-	1
Oak	2	-	1	2	1	-

T_2

	LG			HG		
	C	R	P	C	R	P
Garland	1	1	1	1	1	2
Neepawa	3	1	2	1	2	3
Plum	1	1	1	1	-	-
Oak	2	-	1	2	1	1

T_3

	LG			HG		
	C	R	P	C	R	P
Garland	1	1	1	2	2	4
Neepawa	1	1	2	-	1	3
Plum	1	1	1	-	-	1
Oak	2	-	1	2	1	1

T_4

	LG			HG		
	C	R	P	C	R	P
Garland	1	1	1	2	2	4
Neepawa	2	1	2	1	2	3
Plum	1	1	1	-	-	1
Oak	2	-	1	2	1	1

The result is not, however, efficient (optimally precise) because the means should be average using some weighted average that takes into account their lack of equal precision (due to lack of balance). Also, in a non-orthonogonal experiment, it is difficult to obtain confidence intervals for means or mean differences.