Relation of Physical and Chemical Characteristics of Streams to Fish Communities in the Red River of the North Basin, Minnesota and North Dakota, 1993-95

By Robert M. Goldstein, Joseph C. Stauffer, Peter R. Larson, and David L. Lorenz

U.S. Geological Survey
Water-Resources Investigations Report 96-4227

Contribution from the National Water-Quality Assessment Program

Mounds View, Minnesota
1996

U.S. DEPARTMENT OF THE INTERIOR

BRUCE BABBITT, Secretary

U.S. GEOLOGICAL SURVEY

Gordon P. Eaton, Director

For additional information write to:

District Chief
U.S. Geological Survey
2280 Woodale Drive
Mounds View, MN 55112

Copies of this report can be purchased from:

U.S. Geological Survey
Branch of Information Services
Box 25286
Denver, CO 80225

FOREWORD

The mission of the U.S. Geological Survey (USGS) is to assess the quantity and quality of the earth resources of the Nation and to provide information that will assist resource managers and policymakers at Federal, State, and local levels in making sound decisions. Assessment of water-quality conditions and trends is an important part of this overall mission.

One of the greatest challenges faced by water-resources scientists is acquiring reliable information that will guide the use and protection of the Nation's water resources. That challenge is being addressed by Federal, State, interstate, and local water-resource agencies and by many academic institutions. These organizations are collecting water-quality data for a host of purposes that include: compliance with permits and water-supply standards; development of remediation plans for a specific contamination problem; operational decisions on industrial, wastewater, or water-supply facilities; and research on factors that affect water quality. An additional need for water-quality information is to provide a basis on which regional and national-level policy decisions can be based. Wise decisions must be based on sound information. As a society we need to know whether certain types of water-quality problems are isolated or ubiquitous, whether there are significant differences in conditions among regions, whether the conditions are changing over time, and why these conditions change from place to place and over time. The information can be used to help determine the efficacy of existing water-quality policies and to help analysts determine the need for, and likely consequences, of new policies.

To address these needs, the Congress appropriated funds in 1986 for the USGS to begin a pilot program in seven project areas to develop and refine the National Water-Quality Assessment (NAWQA) Program. In 1991, the USGS began full implementation of the program. The NAWQA Program builds upon an existing base of water-quality studies of the USGS, as well as those of other Federal, State, and local agencies. The objectives of the NAWQA Program are to:

- Describe current water-quality conditions for a large part of the Nation's freshwater streams, rivers, and aquifers.
- Describe how water quality is changing over time.
- Improve understanding of the primary natural and human factors that affect water-quality conditions.

This information will help support the development and evaluation of management, regulatory, and monitoring decisions by other Federal, State, and local agencies to protect, use, and enhance water resources.

The goals of the NAWQA Program are being achieved through investigations of 60 of the Nation's most important river basins and aquifer systems, which are referred to as study units. These study units are distributed throughout the Nation and cover a diversity of hydrogeologic settings. More than two-thirds of the Nation's freshwater use occurs within the 60 study units and more than two-thirds of the people served by public water-supply systems live within their boundaries.

National synthesis of data analysis, based on aggregation of comparable information obtained from the study units, is a major component of the program. This effort focuses on selected water-quality topics using nationally consistent information. Comparative studies will explain differences and similarities in observed water-quality conditions among study areas and will identify changes and trends and their causes. The first topics addressed by the national synthesis are pesticides, nutrients, volatile organic compounds, and aquatic biology. Discussions on these and other water-quality topics will be published in periodic summaries of the quality of the Nation's ground and surface water as the information becomes available.

This report is an element of the comprehensive body of information developed as part of the NAWQA Program. The program depends heavily on the advice, cooperation, and information from many Federal, State, interstate, Tribal, and local agencies and the public. The assistance and suggestions of all are greatly appreciated.

Robert M. Hirsch
Chief Hydrologist

Contents

Illustrations

Illustrations—continued

Tables

Tables—continued

Conversion Factors, Vertical Datum, and Abbreviated Water-Quality Units

Multiply	By	To obtain
centimeter (cm)	0.3937	inch
meter (m)	3.281	foot
kilometer (km)	.6214	mile
square kilometer (km^2)	.3861	square mile
cubic meter per second (m^3/s)	35.31	cubic foot per second
degree Celsius (°C)	1.8 (°C) + 32	degree Fahrenheit (°F)

Chemical concentrations are given in metric units. Chemical concentrations of substances in water are given in milligrams per liter (mg/L) or micrograms per liter (µg/L). Milligrams per liter is a unit expressing the concentration of chemical constituents in solution as weight (milligrams) of solute per unit volume (liter) of water. One thousand micrograms per liter is equivalent to one milligram per liter. For concentrations less than 7,000 mg/L, the numerical value is the same as for concentrations in parts per million.

Sea level In this report "sea level" refers to the National Geodetic Vertical Datum of 1929—a geodetic datum derived from a general adjustment of the first-order level nets of both the United States and Canada, formerly called "Sea Level Datum of 1929".

Use of brand names in this report is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

Relation of Physical and Chemical Characteristics of Streams to Fish Communities in the Red River of the North Basin, Minnesota and North Dakota, 1993-95

By Robert M. Goldstein, Joseph C. Stauffer, Peter R. Larson, and David L. Lorenz

Abstract

Fish community composition was determined at 33 reaches (average length 150 meters) at 22 sites in the Red River of the North Basin during 1994. Sites were selected to represent a range of stream sizes and ecoregions within the basin. Physical and chemical characteristics (classified in data sets of instream habitat, terrestrial habitat, hydrology, and water quality) were determined for various sites for periods ranging from two to 48 years. Instream habitat measurements were made from 1993 through 1995 for 31 reaches at 19 sites. Terrestrial habitat measures of land use/land cover, soils, and riparian zones were determined from a geographical information system coverage for 23 reaches at 14 sites. The geographical information system coverage used data from aerial photographs taken from 1990 and 1991, National Wetlands Inventory data, soils maps from the Natural Resource Conservation Service, and data from the U.S. Department of Agriculture soils data base. Water chemistry data were collected from 14 sites in the basin from 1993 through 1994. Hydrologic variability was determined from U.S. Geological Survey gaging records. Correlation analysis, cluster analysis, and principal components analysis were used to determine representative variables which accounted for the most variation in each data set. The representative variables and the fish community data were analyzed with canonical correspondence analysis to determine the relative effect of each source of environmental influence on fish community composition. Instream habitat, terrestrial habitat, and hydrologic variability were analyzed together. Water chemistry data were analyzed separately due to a lack of corresponding sites.

Within the instream habitat data set, measures of habitat volume (channel width and depth) and habitat diversity were most significant in explaining the variability of the fish communities. The amount of nonagricultural land and riparian zone integrity from the terrestrial habitat data set were also useful in explaining fish community composition. Variability of mean monthly discharge and the frequency of high and low discharge events during the three years prior to fish sampling were the most influential of the hydrologic variables.The first two axes of the canonical correspondence analysis accounted for 43.3 percent of the variation in the fish community and 52.5 percent of the variation in the environmental-species relation. Water-quality indicators such as the percent of fine material in suspended sediment, minimum dissolved oxygen concentrations, minimum concentrations of dissolved organic carbon, and the range of concentrations of major ions and nutrients were the variables that were most important in the canonical correspondence analysis of water-quality data with fish. No single environmental variable or data set appeared to be more important than another in explaining variation in the fish community. The environmental factors affecting the fish communities of the Red River of the North are interrelated. For the most part, instream environmental conditions (instream habitat, hydrology, and water chemistry) appear to be more important in explaining variability in fish community composition than factors related to the agricultural nature of the basin.

Introduction

The U. S. Geological Survey (USGS) National Water Quality Assessment (NAWQA) Program was initiated to define the current status and trends in the quality of the nation's surface and ground-water resources. Because the amount and geographical distribution of these resources are so vast, the major activities of NAWQA will occur within 60 hydrologic systems (study units) across the country accounting for about 70 percent of the nation's water use and population served by public water supply. The implementation plan (Leahy and others, 1990) specified 20 study units to be operational during each of three cycles of NAWQA. Each cycle is to include three years of intensive study followed by six years of low-intensity monitoring. Cycles are to be initiated at three-year intervals.

In 1991, the USGS began to implement the field studies of the first cycle of NAWQA in the Red River of the North (herein referred to as the Red River) Basin and 19 other study units across the country. The goals of the program (Cohen and others, 1988) are to: (1) Provide a nationally consistent description of current water-quality conditions for a large part of the nation's water resources; (2) Define long-term trends (or lack thereof) in water quality; (3) Identify, describe, and explain the major factors that affect observed water-quality conditions and trends. These goals were established so the program would not only define the current conditions, but also determine the reasons for observed conditions and monitor long-term trends. NAWQA is using a multidisciplinary approach to assess water quality. The ecology of aquatic biological communities is one of the disciplines that is to be used to provide multiple lines of evidence for the assessment (Gurtz, 1994).

The Red River Basin was selected as a study unit because it represents an important hydrologic region where good water quality is vital to the region's economy, and the water quality of the Red River, which flows northward into Manitoba, Canada, is of international concern. The Red River Basin is an important agricultural area, and the northern location of the Red River Basin and potential interaction of surface and ground water are essential factors for a complete national assessment of water quality (Stoner, 1991).

Part of this study addresses the relation of physical and chemical characteristics of the Red River Basin to the fish communities of its streams. The emphasis of this effort was to determine whether natural or anthropogenic factors explained the variability of fish communities in the basin.

Purpose and Scope

The purpose of this report is to assess the composition of the fish communities of streams in the Red River Basin by determining the relation of community composition to surrounding landscape features, physical habitat, hydrology, and water quality. The report includes information from the Red River Basin in the United States on physical habitat of streams, hydrologic variability, water quality, and landscape features such as land use, land cover, and soils. Data sources are from measurements and collections made from 1992 through 1995 and the hydrologic record available from 1947 through 1994. These analyses are part of the multiple lines of evidence NAWQA uses to assess aquatic resource quality in the Red River Basin.

Study Area Description

The study area is the Red River Basin in the United States (fig. 1). The basin includes those portions of the Roseau and Pembina River Basins within the United States. The Red River Basin covers over 90,000 km^2; this area does not include the noncontributing Devils Lake Basin in North Dakota. The Red River of the North Basin is located near the geographic center of North America.

The Red River Basin has little topographic relief, approximately 490 m. Surface elevations range from 716 m above sea level at the far western edge of the basin in North Dakota to 228 m above sea level where the Red River crosses into Canada. From its formation at the confluence of the Bois de Sioux and Otter Tail Rivers at Wahpeton, North Dakota and Breckenridge, Minnesota, the river flows northward to the Canadian border. The elevation of the river decreases 61 m over its 640 km course to the border. The Red River Basin contains four major physiographic areas (Maclay and others, 1972; Winter and others, 1984): (A) the Drift Prairie, (B) the Red River Valley Lake Plain, (C) the Lake-Washed Till Plain, and (D) the Moraine (fig. 2). The physiographic areas are closely related to the major U.S. Environmental Protection Agency ecoregions (Omernik and Gallant, 1988) in the basin. Stoner and others (1993) have described the environmental setting of the Red River Basin in detail.

Ecoregions

Ecoregions are "regions of relative homogeneity in ecological systems or in relationships between organisms and their environment" (Omernik and Gallant, 1988, p. 1). Classification of ecoregions is based on land use, potential natural vegetation, land-surface form, soils, climate, and length of growing season (Omernik and Gallant, 1988). Within the Red River Basin (fig. 2) there are six ecoregions. From west to east they are: (1) Northwestern Glaciated Plains, (2) Northern Glaciated Plains, (3) Red River Valley, (4) North Central Hardwood Forests, (5) Northern Lakes and Forests, and (6) Northern Minnesota Wetlands. The Northwestern Glaciated Plains is a minor area on the far western border of the basin and so is included with the Northern Glaciated Plains in figure 2.

The Northern Glaciated Plains ecoregion is almost entirely within North Dakota with a small portion

Figure 1. Location of the Red River of the North Basin study unit.

extending into Minnesota at the southeastern edge of the basin. This region is rolling hills with many small, closed basin ponds called prairie potholes. Many of the small streams in this ecoregion are intermittent, and cease to flow during the summer (Stoner and others, 1993). The Red River Valley ecoregion is the bed of Glacial Lake Agassiz. The streams in this ecoregion are very low gradient, meandering streams (Stoner and others, 1993; Goldstein, 1995). The Minnesota side of the basin contains the other three ecoregions. The North Central Hardwood Forests are a transition zone between the Red River Valley and the Northern Lakes and Forests ecoregion (Stoner and others, 1993), and the Northern Minnesota Wetlands is a large, flat expanse of wetlands, peat bogs, and marshes. Stoner and others (1993) have characterized the ecoregions of the Red River Basin. Fandrei and others (1988) have presented details of water quality and land use for ecoregions in Minnesota.

Fish communities in streams of the Red River Basin differ not only between ecoregions but also with the number of ecoregions the stream flows through. A cluster analysis of fish species from rivers and streams of the Red River Basin identified five river cluster groups (Goldstein, 1995). Three of the clusters corresponded to ecoregions while the other two clusters reflected the number of ecoregions the rivers flowed through. As the number of ecoregions a river flowed through increased, the species richness (number of species) also increased (Goldstein, 1995). The increase in species richness was attributed to greater habitat diversity and organic energy sources available as streams traversed the increased number and types of ecoregions (Goldstein, 1995).

Physical and chemical characteristics of the basin

The Red River Basin is predominantly agricultural (fig 3). Sixty-four percent of the land is cropland. When pasture is added to the cultivated land the percentage increases to over 80 percent (Lorenz and Stoner, 1996). Only 0.56 percent is developed (urban, industrial, or residential), and the remainder is undeveloped land (grassland, wetlands, woodland, water, and other categories). The important crops throughout the basin are small grains, corn, soybeans, edible beans, sunflowers, sugar beets, and potatoes. Although small grains are predominant, crops of corn and soybeans are locally dominant in the southern one-third of the basin. Potatoes and edible beans, which are rotated with wheat, are important secondary crops in the northern one-half of the basin (Stoner and others, 1993). The Red River Valley is the most agricultural ecoregion in the study unit. To the west, in the Northern Glaciated Plains, there is more pasture, hay, and sunflowers. In the eastern ecoregions, the North Central Hardwood Forests and the Northern Lakes and Forests, the amount of woodland and pasture also increases. In the Northern Minnesota Wetlands there is some agriculture, mostly small grains, hay, and pasture, but natural wetlands are predominant.

There are eight major soil associations in the basin which support the extensive agriculture (Lorenz and Stoner, 1996; Omodt and others, 1968; Minnesota Soil Survey Staff, 1983; U.S. Department of Agriculture, 1975, 1977a and 1977b). Soil associations are based on similarities of slope, texture, natural drainage, and special features. Land-use and cropping patterns are linked to these soil associations. The western portion of the basin contains three associations: (1) hilly steep land with thin erodible soils, mostly ustolls (dry prairie soils) and borolls (northern prairie soils); (2) black, limey, clayey soils, which are aquoils (wet soils due to retained

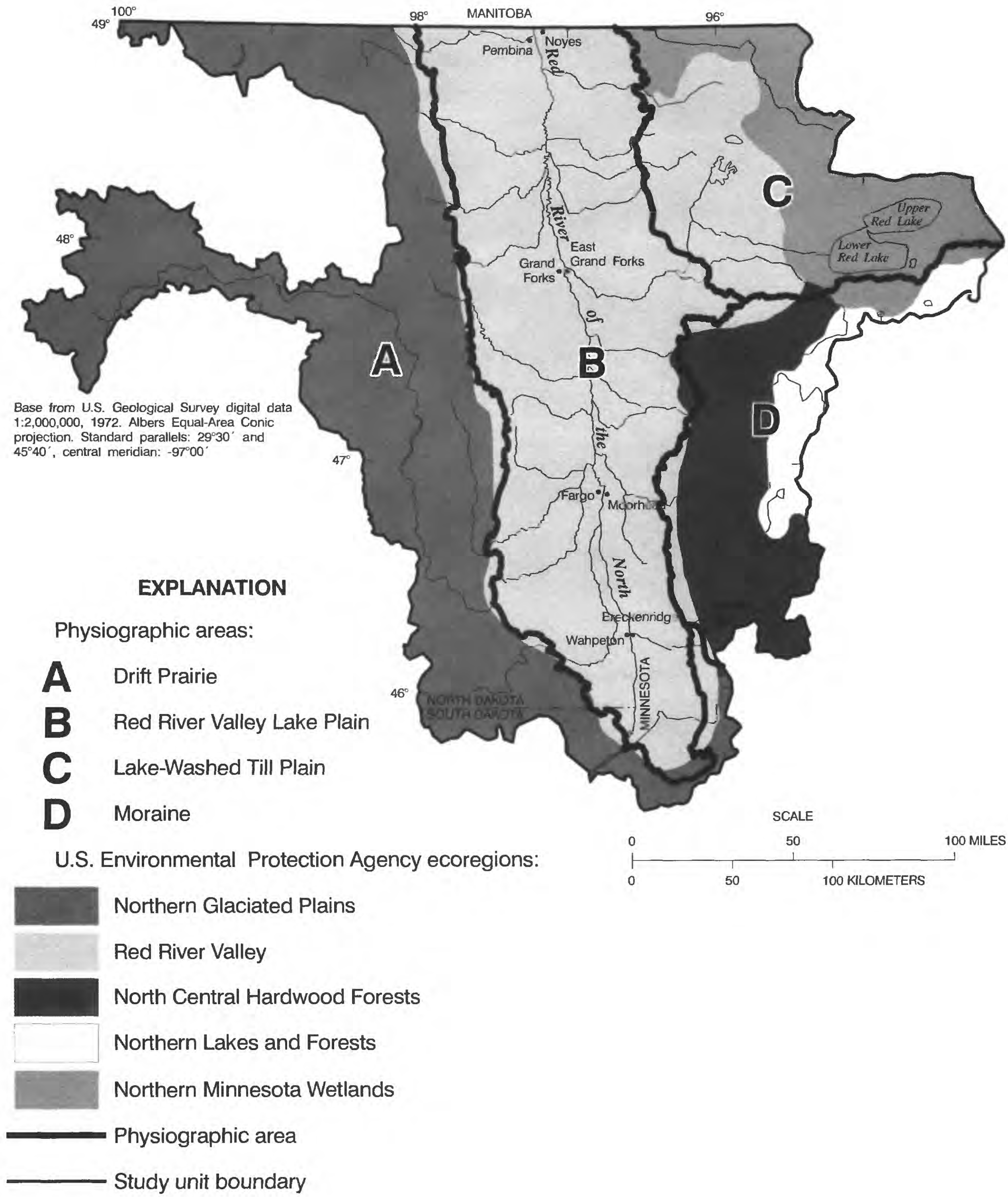

Figure 2. Ecoregions (modified from Stoner and others, 1993) and physiographic areas (from Maclay and others, 1972; Winter and others, 1984) in the Red River of the North Basin study unit.

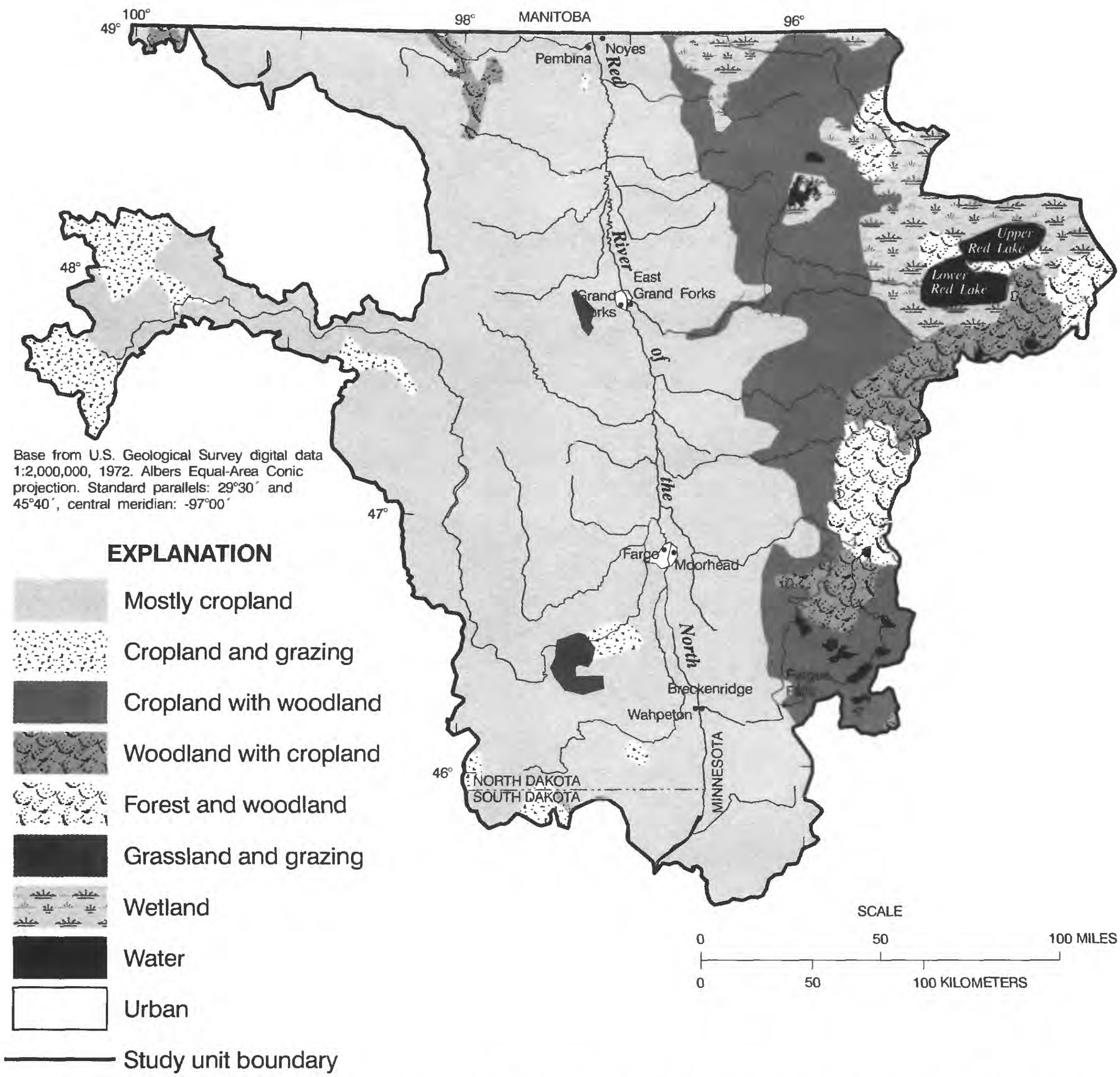

Figure 3. Land use in the Red River of the North Basin study unit (from Lorenz and Stoner, 1996).

spring moisture); and (3) black, loamy soils. The central portion of the basin which is almost entirely agricultural, also contains black, limey, clayey soils. In addition the central portion of the basin contains clayey soils and sandy soils. The eastern part of the basin contains four soil associations: (1) organic soils, (2) black, limey, clayey soils, (3) loamy soils, and (4) gray soils and black loams in the wooded areas. Additional information on soils can be found in Stoner and others (1993).

Mean annual precipitation increases from about 43 cm in the western part of the basin to 66 cm in the east. Approximately 75 percent of the annual precipitation occurs from April through September, while December through February are usually the driest months (Stoner and others, 1993). Precipitation varies greatly between wet and dry periods that range from a year to a decade. The most recent dry period was 1988-1990 (Stoner and others, 1993). The ratio of the lowest mean decadal streamflow to the highest (ten years ending 1940 and 1975, respectively, measured at Grand Forks, North Dakota) was 7.5. This ratio is indicative of the hydrologic variability in this basin.

Hydrologic data on stream and river discharges, which includes historical data from the period of record, are reported in Mitton and others (1994) and Harkness and others (1994). The Red River receives most of its flow from the eastern tributaries because of regional patterns of precipitation, evapotranspiration, soils, and topography (Stoner and others, 1993). Annual streamflow varies greatly with most streamflow occurring in spring and early summer due to snowmelt. About one-half the average annual streamflow of the Red River in the United States crosses the Canadian border during April and May, and 70 percent from April through July (Stoner and others, 1993). Although most of the streamflow from the North Dakota side of the basin occurs during April and May, that is not the case on the Minnesota side where lakes and wetlands moderate runoff.

The hydrologic regime in the basin varies through the year. In the spring, the water temperature increases and begins to melt ice, which has completely covered the rivers. As the water level and flow increase, the ice cover is broken and begins to move downstream. Historically, floods on the Red River were due to the extremely low gradient of the streams, and in the early spring to ice jams preventing spring meltwater from the southern part of the basin from flowing downstream to the northern and still frozen part of the basin. Annual high discharges follow ice-out as meltwater runs off the still frozen ground. Much of the Red River Basin has been ditched to improve drainage (Waters, 1977). An extensive network of drainage ditches may exacerbate the situation by increasing the rate of runoff (Goldstein, 1995). Streamflow decreases through spring and summer, but can be periodically increased by heavy rainfall. Some of the streams exhibit rapid increases in flow and water levels may rise several feet in one day. Low flow or annual minimum flow is usually reached in late summer or early fall.

During 1993 through 1995, the hydrologic regime in the Red River was unusual compared to the previous years. The basin experienced three consecutive years of record-high precipitation and high stream discharges. The annual mean discharge of the Red River at the Canadian border from 1988 to 1992 was 54 m^3/s. In 1993, 1994, and 1995 the annual mean discharges were 185, 186, and 240 m^3/s, respectively. Additional information on hydrology and climate are available in Stoner and others (1993).

Water quality in the Red River Basin has been summarized for selected constituents by Maclay and others (1972) for Minnesota and Winter and others (1984) for North Dakota. Tornes and Brigham (1994) have summarized certain aspects of water quality in the Red River Basin from 1970 through 1990. Concentrations of dissolved solids in the Red River are generally less than 600 mg/L. Mean annual values range from 347 mg/L near the headwaters to 406 mg/L at the Canadian border (Winter and others, 1984). Calcium and magnesium are the major cations and bicarbonate is the major anion. Dissolved solids are generally lower in the Minnesota tributaries than those in North Dakota (Stoner and others, 1993). Saline seeps in the northern North Dakota portion of the basin may affect stream water quality during low-flow conditions (International Red River Pollution Control Board, 1990). Fluvial sediment concentrations in the Red River (measured by the USGS at the Canadian border) range from 20 mg/L under ice to 270 mg/L during storm runoff. More than 98 percent of the sediment is usually silt size or smaller (Stoner and others, 1993).

Major nutrient concentrations of nitrogen and phosphorus in surface water vary across the basin. Total nitrogen in streams ranges from below detection limit (0.2 mg/L) to greater than 20 mg/L, while dissolved nitrogen ranged from below detection (0.2 mg/L) to 3.4 mg/L and averaged almost 80 percent of the total nitrogen (Tornes and Brigham, 1994). Approximately two-thirds of the nitrogen in the rivers is total organic nitrogen and ammonia. Total phosphorus concentrations range from approximately 0.05 mg/L to 1.4 mg/L at USGS sampling sites (Tornes and Brigham, 1994). The higher nutrient concentrations have been attributed to fertilizer use and sewage treatment discharges based on the seasonal occurrence of the higher concentrations in the spring during runoff and localized higher concentrations downstream of major urban areas with sewage treatment (Tornes and Brigham, 1994).

Water quality varies by ecoregion. Fandrei and others (1988) compared water quality among ecoregions in Minnesota. Although their study included ecoregions entirely within the Red River Basin, most of the ecoregions extend beyond basin boundaries. They found that conductivity, total suspended solids, nitrite plus nitrate concentrations, fecal coliform bacteria, and turbidity generally decreased from the Northern Glaciated Plains through the Red River Valley to the North Central Hardwood Forest and were lowest in the Minnesota Wetlands and Northern Lakes and Forests. Total phosphorus and ammonia nitrogen were highest in the Red River Valley; whereas, pH was lowest in the Northern Lakes and Forests.

About 14,000 years before present, the Wisconsin stage glacier covered the basin (Wright, 1972). Much of the basin was covered by Glacial Lake Agassiz for approximately 4,000 years. The series of ice advances and retreats combined with the various aquatic connections to glacial period biotic refugia account for the biological communities inhabiting the basin today. By approximately 8,500 years before present, the outlets had stabilized leaving the Red River Basin and drainage patterns observed today (Underhill, 1989). Underhill (1989) contends that almost all fish species found in the Red River Basin today are the result of immigration from the Mississippi River during periods of connections with the Red River. These communities have been modified by cultural activities (agriculture) and hydraulic alterations (dams, channelization of streams, draining of wetlands, and ditching) in the basin during the last 150 years of European settlement and development. The fish communities of the rivers and streams in the basin are not static. Fish community composition of the streams changes both spatially and temporally with the changing environmental conditions (Goldstein, 1995).

Acknowledgments

The authors recognize the efforts of the entire Red River of the North NAWQA team and the many USGS people who contributed to the collection of data and analyses. Tara Williams-Sether provided hydrologic statistics. Gary Burkhart and Kelly Boespflug collected many of the water-quality samples under less than ideal conditions. The report has been improved through discussions with Kathy Lee, Stephan Porter, J. Bruce Moring, Michael Meador, Thomas Cuffney, and John Kingston. The authors acknowledge the guidance provided by the members of the Red River of the North Liaison Committee and the land owners of the basin who allowed access to sampling sites.

Methods

This section describes the site selection process, measurements of physical and chemical characteristics of sites, fish collection, and a summary of data analysis. Details of data analysis and the selection of specific physical and chemical variables are presented in the "Supplemental Information" section.

Site Selection

Sampling sites were selected to represent important environmental settings in the study unit (Gurtz, 1994). A multidisciplinary team, which included a hydrologist, hydrogeologist, water-quality specialist, environmental chemist, geographic information systems (GIS) specialist, and an aquatic biologist, helped select sampling sites. The basin was stratified by applying several criteria. The first level of stratification was by ecoregion or physiographic area. The second level of stratification was by land use or dominant cropping pattern within each ecoregion. The third level of stratification was by basin size and hydrologic contribution. Because of the multidisciplinary nature of the NAWQA Program and the historical data base on hydrology and water chemistry, efforts were made to select sites at existing USGS stream gages.

The site selection process involved three phases. The first phase was conceptual and involved examination of maps to select candidate sites based on locations of ecoregion, gaging-stations, and roadway access. The second phase was a field reconnaissance of potential sites. The objective of the reconnaissance survey was to visit each potential site and determine (1) site accessibility, (2) general aquatic habitat conditions, (3) proximity of potential natural or anthropogenic sources of degradation, (4) conditions for sampling, (5) potential sampling equipment needs and methods, and (6) the overall representativeness of the site compared to other sites in the same site selection stratum. The third phase of the site selection process involved evaluating each of the 114 sites visited during the second phase and selecting 30 sites for further investigation with the goal of determining 12-15 sites for study.

During August 1992, a preliminary ecological survey was conducted at the 20 wadeable sites of the 30 sites identified by the site selection process. Nonwadeable sites were not included in this survey. The purpose of the survey was to test equipment and methods of collection as well as to determine some physical and biological community parameters (channel condition, number and types of geomorphological units, stream stage, bank slopes, bank stability, erosion, riparian cover, substrate types, instream cover, and relative abundance of macroinvertebrates and fish) for further study.

Fish were collected from wadeable sites by electrofishing with a backpack electrofishing unit (Smith-Root Model 12 B). A single collection pass was made that emphasized collecting in all available habitat types. The numbers of species collected in a limited time period from small-size and medium-size wadeable streams were compared. A nonparametric test (Kruskal-Wallis, $p<0.05$) indicated that species richness was different between small and medium-sized streams. A nonparametric test was used because of the conservative nature of the test and limited level of comparability of effort among sites. The measure of stream size used in the analysis was stream order (Strahler, 1957) as determined from USGS 1:100,000- scale maps. Small streams were third order or less (mean channel width < 8 m) and medium streams were fourth order (mean channel width 8-15 m). Small streams had fewer species than the medium size streams. The increase in species richness with increasing stream size has been documented in other basins (Shelford, 1911; Sheldon, 1968; Goldstein, 1981) and was the reason for conducting this analysis prior to site selection. This preliminary sampling and data analysis indicated the need for stratifying ecological sampling sites by stream size within ecoregions. Because the hydrologic conditions vary across the basin, the Red River Valley ecoregion was separated into two halves. The Red River Valley West contains the streams subject to intrusion of saline ground water during low-flow conditions (Stoner and others, 1993). Annual discharge was standardized by basin area to determine annual runoff. There was a significant difference (Wilcoxon rank sum test, $p<0.008$) in the average annual runoff between the western saline (mean of 3.3 cm) and eastern (mean of 8.8 cm) parts of the Red River Valley ecoregion.

Based on reconnaissance, preliminary sampling, and input from discipline specialists, 22 sites were selected for biological community and habitat sampling (fig. 4, table 1). Allocation of the number of sites by ecoregion and strata was based on (1) the relative size of the ecoregion to the entire basin, and (2) the hydrologic contribution of the basin and river to the annual discharge of the Red River at the Canadian border. Fifteen sites were selected for water-quality sampling. Water-quality parameters were determined at each of the sites, with the addition of pesticide determinations at four of the sites. Some of the water-quality sites were not suitable for ecological sampling due to problems with access, nearby dams, or other factors. There were eleven sites common to both ecological sampling and water-quality sampling.

Ecological sites were further defined as either intensive ecological sites or limited ecological sites. At each intensive ecological site three reaches were established for sampling; whereas, one reach was defined at the limited sites. Stream sites were also classified by drainage area upstream of the site (stream order proved difficult to use due to the differences in precipitation across the basin) and whether the stream was wadeable or not. The six intensive ecological sites were located in each major ecoregion and at the Canadian border. Two of the six intensive ecological sites were located on the eastern and western sides of the Red River Valley ecoregion (the Red River was the boundary) because (1) there is a precipitation gradient of increasing annual precipitation from west to east (Stoner and others, 1993), (2) the northwestern streams are subject to intrusion of saline ground water (Lorenz and Stoner, 1993), and (3) the ecoregion accounts for almost half of the area in the basin.

Four sampling sites were classified as large nonwadeable, two of which were on the mainstem of the Red River. Drainage basin sizes of large rivers were greater than 10,000 km^2 and runoff (1993-94) greater than 5.0 cm per year except for the Roseau River, which had a drainage area about 4,000 km^2, but had more than 7.5 cm of runoff per year and was not wadeable. The presence of large impoundments upstream of the Roseau River site contributed to a more regulated flow and greater water depths. Nine sites were medium-size wadeable streams. One site on the Bois de Sioux River was considered a mainstream site because it was approximately 15 km upstream from the formation of the Red River by the confluence of the Bois de Sioux and Otter Tail Rivers, and there were no impediments to fish movement between that site and the Red River. Drainage areas of medium-size wadeable streams were between 3,200 km^2 and 10,000 km^2. Runoff in medium streams on the western side of the basin was less than 3.8 cm per year during 1993 and 1994, and more than 7.6 cm per year on the eastern side of the basin. One exception was the Otter Tail River, which at the sampling site at Perham, has a drainage area of only 870 km^2. The Otter Tail River flows through a series of lakes and ponds, which maintain the river at a channel width and depth more comparable to medium-size streams than to small streams. Downstream of the ecological sampling site, at a water-quality sampling site, the mean annual runoff of the Otter Tail River was 12.3 cm, which is more than any of the other streams sampled in the basin. The remaining nine sites were on small-size wadeable streams. Drainage areas of small streams were less than 3,200 km^2 (table 1).

Site Characterization

Physical and chemical characteristics of each sampling site were quantified from 1992 through 1995. The physical and chemical characteristics were separated into four data sets: instream habitat, terrestrial habitat, hydrology, and water quality. These data sets are the environmental data sets.

Habitat

Quantification of habitat features was conducted at multiple spatial scales of stream reach, segment, and basin (Meador and others, 1993a). This spatially hierarchical approach is based on a a system of stream organization (Frissell and others, 1986; Lotspeich and Platts, 1982) and geomorphology. For purposes of this study, reach-level habitat features are referred to as

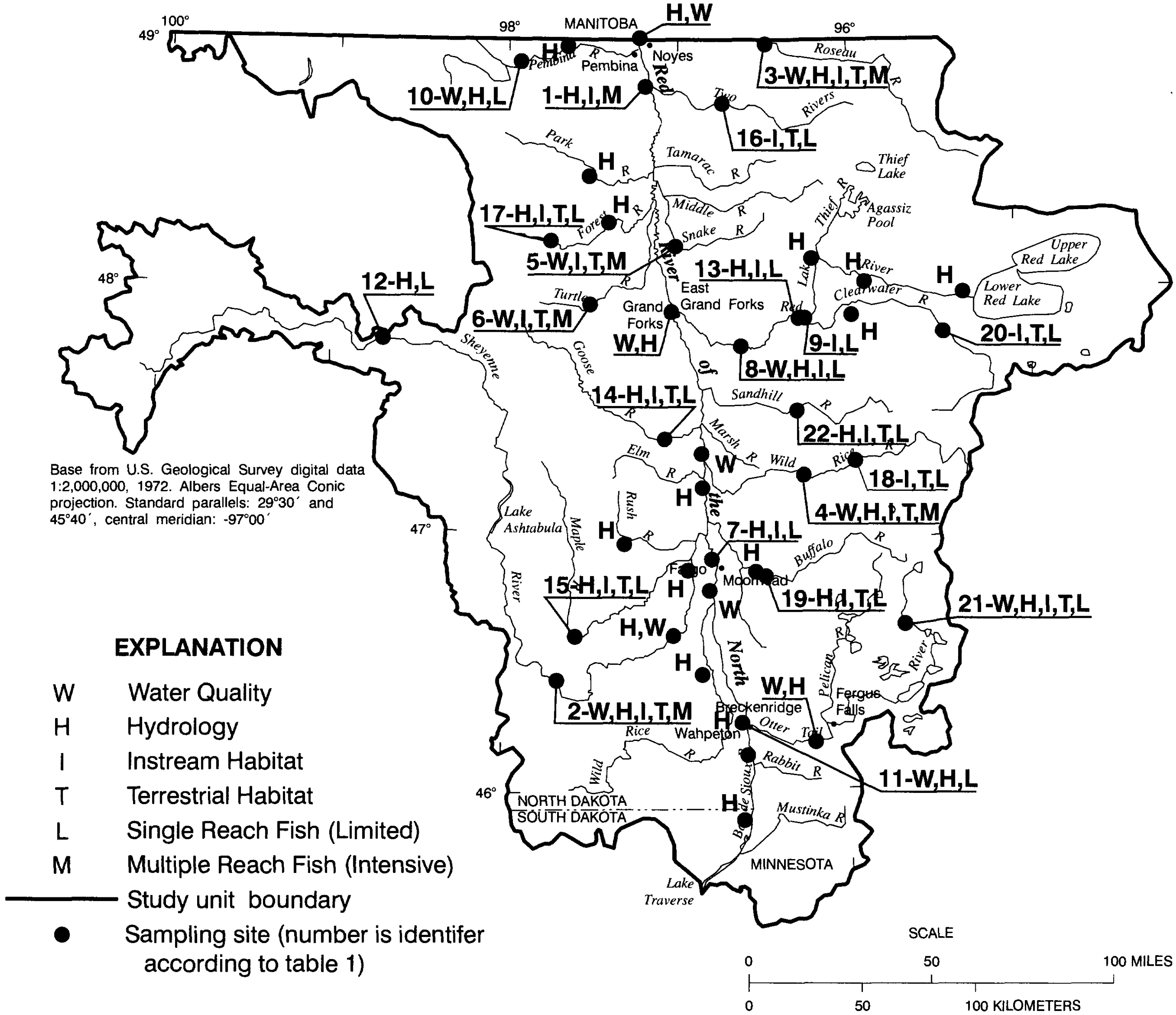

Figure 4. Sampling sites in the Red River of the North Basin study unit.

instream habitat and include the stream banks, geomorphological, and wetted features of the stream. Segment and basin features and characteristics are referred to as terrestrial habitat to denote the dependance of aquatic resources on the surrounding terrestrial environment and include aspects of the riparian zone, anthropogenic features such as dams and ditches, soil characteristics, and land use/land cover.

Instream habitat

A reach is the length of stream sampled. A reach was determined in one of three ways in the following priority: (1) two replications of two of three different geomorphological units (for example, two riffles and two pools, or two runs and two riffles), (2) a distance equal to 20 times the width of the stream for meandering streams, or (3) a minimum distance criterion of at least 150 m for wadeable streams and 500 m for nonwadeable streams (Meador and others, 1993a). Maximum reach length criteria were 300 m for wadeable streams and 1,000 m for nonwadeable streams. The minimum distance criterion was used at most of the sampling sites. The majority of the streams in the Red River Basin are meandering, prairie streams with small gradient.

Table 1.—Information pertaining to fish sampling sites in the Red River of the North study unit.

[n/a is new gage site, data not available; km^2, square kilometers]

Sampling sites	Site number (fig. 4)	Site abbreviations[1]	Ecoregion or Mainstem[2]						Size class and type[3]	Basin area (km^2)	Mean annual runoff (1993-1994) (cm)
			NGP	RVW	RVE	NCLF	NMW	MS			
Intensive sites											
Red River at Joliette, N.D.	1	Jol						X	L/ NW	92,200	6.0
Sheyenne River at Lisbon, N.D.	2	Shey-lisbon	X						M /W	11,600	3.1
Roseau River at Caribou, Minn	3	Roseau					X		L/ NW	4,020	8.5
Wild Rice River at Twin Valley, Minn.	4	Wrtv				X			M/ W	2,400	11.8
Snake River at Alvarado, Minn.	5	Snake			X				S/W	658	7.0
Turtle River at Turtle River State Park	6	Turtle		X					S /W	565	5.2
Limited sites											
Red River at Fargo, N.D.	7	Fargo						X	L/ NW	17,300	8.4
Red Lake River at Crookston, Minn.	8	Rlrcrks			X				L/ NW	13,650	8.9
Red Lake River at Red Lake Falls, Minn.	9	Rlrrlf			X				M/ W	9,270	n/a
Pembina River at Walhalla, N.D.	10	Pembina		X					M /W	8,290	2.9
Bois de Sioux River at Doran, Minn.	11	Bsdesx						X	M /W	4,870	10.5
Sheyenne River at Warwick, N.D.	12	Shey-warwick	X						M /W	4,860	1.5
Clearwater River at Red Lake Falls, Minn.	13	Cwrlf			X				M /W	3,520	8.7
Goose River at Hillsboro, N.D.	14	Goose		X					M /W	3,210	2.4
Maple River at Enderlin, N.D.	15	Maple	X						S /W	2,140	5.3

Table 1.—Information pertaining to fish sampling sites in the Red River of the North study unit—Continued.

Sampling sites	Site number (fig. 4)	Site abbrev-iations [1]	Ecoregion or mainstem[2] NGP	RVW	RVE	NCLF	NMW	MS	Size class and type[3]	Basin area (km^2)	Mean annual runoff (1993-1994) (cm)
Two Rivers at Lake Bronson, Minn.	16	Tworiv			X				S /W	1,150	n/a
Forest River at Fordville, N.D.	17	Forest		X					S /W	1,150	2.6
Wild Rice River at Mahnomen, Minn.	18	Wrmah				X			S /W	1,000	n/a
Buffalo River at Buffalo River State Park, Minn.	19	Buffalo			X				S /W	980	11.2
Clearwater River at Berner, Minn.	20	Cwbern					X		S /W	930	n/a
Otter Tail River at Perham, Minn.	21	Otter				X			M /W	870	n/a
Sandhill River at Fertile, Minn.	22	Sand			X				S/W	560	8.3

[1] Intensive sites have three reaches identified by suffix 1, 2, or 3.
[2] NGP=Northern Glaciated Plains, RVW=Red River Valley West (includes saline intrusion area in northwestern tributaries), RVE=Red River Valley East, NCLF=North Central Hardwood Forests and Northern Lakes and Forests, NMW=Northern Minnesota Wetlands, and MS=Mainstem.
[3] L=Large river, M=Medium river, S=Small river NW=Nonwadeable, W=Wadeable.

Replication of geomorphological units was not evident at most sites which are dominated by the presence of runs. In some cases, runs were the only geomorphological unit present and the criterion of 20 times the stream width was not used because it resulted in reach lengths in excess of the maximum length criteria.

At each reach a reference location was established. The reference location is a permanent structure or feature such as a USGS gage, bridge, building, or geographical marker. Six equally spaced transects were established along the length of the reach for all wadeable reaches. Three equally spaced transects were established at nonwadeable reaches. Physical measurements made at each transect related to the flood plain, bank features, riparian vegetation, instream features, substrates, geomorphological units, and overhead canopy (Meador and others, 1993a). Instream measurements of depth, velocity, substrate, embeddedness (if applicable), and cover for fish were made at three points equally spaced across the channel along each transect. Additionally, a diagram of the reach was made that included the locations of the transects, surrounding land use, and major habitat and geomorphological features. For details of the habitat measurement protocols see Meador and others (1993a).

Habitat measurements were made as early in the year as possible to avoid vegetation obscuring bank, riparian, and floodplain features. Physical habitat measurements were made at all three reaches at the multiple reach sites during April, May, and June 1993. Habitat measurements were made at the single reach sites during April, May, and June 1994. High water prevented habitat measurements at three sites. All measurements were made by the same person for consistency.

Terrestrial habitat

A stream segment is the part of a stream which includes the sampling reach and is bounded by tributary inflows or major physical or chemical discontinuities such as abrupt changes in gradient or a point source discharge (Frissell and others, 1986). In this analysis, a stream segment is arbitrarily defined as the area a specific distance upstream from the upper boundary of the reach and laterally from the reach on each side of the stream (the distance was based on the area, inclusive of the reach, readily visible on aerial photographs, Supplemental Information section). Segments are considered to have relatively homogeneous physical, chemical, and biological attributes (Frissell and others, 1986; Meador and others, 1993a).

Terrestrial habitat data were obtained from several sources. Soil series maps were obtained from the Natural Resource Conservation Service (NRCS) for each county (if available) containing a sampling site (U.S. Department of Agriculture, 1975, 1977a, 1977b). Aerial photographs of each site were obtained from the National Aerial Photography Program (NAPP) of the USGS which surveyed North Dakota and Minnesota during 1990 through 1991. Land use/cover data were obtained from the National Wetlands Inventory (NWI) of the NRCS. A vector-based Geographic Information System (GIS) with an Albers equal area projection was selected for data interpretation. The soils maps were scanned and digitized into the GIS. Land-cover data were digitized from the NAPP images, and the NWI data were retrieved as digital line graph files from the NRCS. All maps and photographs were converted to the same Albers equal area projection. Once the combined coverage was completed, terrestrial habitat data were extracted.

Terrestrial habitat data pertaining to soils, land use, stream gradient, and the riparian zone were determined for three set distance classes. These were 300 m, 1,000 m, and 4,000 m measured from the most upstream boundary of the reach and laterally one half the distance on each side of the stream. Using the first distance class (300 m) as an example, at a straight length of stream, this procedure would produce a square area. The dimensions would be 150 m on each side of the stream and 300 m upstream from the upper boundary of the reach. For a very sinuous stream the shape produced would have an outer boundary which followed the course of the stream. Of the three set distance classes, the 4,000 m distance provided the best site discrimination (Supplemental Information section).

Data derived from the GIS coverage included the relative area of each series of soil. Three characteristics of soil were used to define the soils of each distance class: shallow soil permeability, soil erodibility (K factor), and wind erodibility. These data were retrieved from the U.S. Department of Agriculture state soils data base (STATSGO) (U.S. Department of Agriculture, 1991). The range of permeability values was divided into three equal groups and the percent of the total area in each permeability class (low = class 1, moderate = class 2, and high = class 3) was calculated for each group. The potential for erosion by water, soil erodibility, was classified into three groups (low = class 1, moderate = class 2, and high = class 3). The low group had a soil erodibility factor (K) less than 0.2; the moderate group had K from 0.2 to 0.3; and the high group had K greater than 0.3 (Brady, 1984). Wind erodibility is ranked on a scale of 1 to 10, with 1 being most resistant. Three classes were established for wind erodibility: class 1 included ranks 1-3, class 2 included ranks 4-6, and class 3 included ranks greater than 6. The percent of total area was determined for each class of soil characteristic for each of the three distance groups. Also determined from the GIS coverages of the sites were the relative areas of each land use and land cover type, the mean width of the riparian zone, and the coefficient of variation of the width of the riparian zone as measured along six equally spaced transects of the three distance classes. Discontinuity of the riparian zone (both left and right bank) was measured as the linear distance of stream without a riparian buffer divided by the total linear distance of the stream in the distance class. Tree density in the riparian zone was determined for each bank by the point-quarter method (Meador and others, 1993a) and applied to all three distance classes. Sinuosity and stream gradient were determined from USGS 1:24,000-scale topographic maps. The number of tributaries within each distance class was determined from the NWI coverages. Additional data used at the segment and basin level pertained to anthropogenic modifications of streamflow and drainage. These included the number of dams within ten kilometers of the upstream boundary of the reach and the number of dams within ten kilometers downstream of the upstream boundary of the reach. Tributary dams were counted if located within 1 km of the mainstem. Ditches are another anthropogenic factor included in the segment- and basin-level data set. The total linear distance of all drainage ditches in the basin upstream of the site was determined from USGS 1:100,000 scale maps. Complete data were available for 14 sites.

Hydrology

USGS stream gaging records from 33 sites throughout the basin were analyzed (fig. 4). The variables determined from the hydrologic data were expressions of the relative variability and stability of the hydrologic regime. Variables selected for analysis were designed to reflect the predictability of discharge (Colwell, 1974; Horwitz, 1978; Poff and Ward, 1989; Poff and Allen, 1995). These included mean monthly flow, coefficient of variation of mean monthly flow, flood (bankfull stage) frequency, mean annual high and low flows for various time periods ranging from 1 to 90 days, average annual frequency and duration of high and low flow periods (75th and 25th percentiles were selected to indicate high and low flow, respectively), and timing of high and low flow events. Two periods of record were examined. The first was from 1947 through 1994. Even though discharge data were available for most sites for the entire period, some sites had missing data. Therefore, the periods of record varied from 22 to 48 years with most sites having the latter. This period was used to characterize the historical variability of flow. The second period was 1992 through 1994, when hydrologic variability would have the most direct influence on the sampled fish communities in the Red River Basin because the life cycle and longevity of many fish species is only a few years. The long-term data were used to define the variables most important in distinguishing the sites. The short-term data were used in the analysis with the fish community data.

Water quality

General water-quality parameters were measured at selected locations throughout the basin. Some constituents were measured as part of standard USGS monitoring and NASQAN (National Stream Quality Accounting Network). Minnesota data are reported in Mitton and others (1994) and North Dakota data in Harkness and others (1994). Water-quality data were also obtained at locations specifically for NAWQA (Lorenz and Stoner, 1996). Data were derived from samples taken at standard USGS gaging stations, NASQAN stations, and NAWQA sampling sites during 1992 through 1995. As such, the data are representative but not necessarily comprehensive because the number of sites per ecoregion varied as did the periodicity of sampling.

Water-quality constituents measured as part of regular USGS and NASQAN monitoring include specific conductance, pH, temperature, dissolved oxygen, major anions and cations, nutrients (forms of nitrogen and phosphorus), suspended sediment, and dissolved solids (Harkness and others, 1994; Mitton and others, 1994). The NAWQA parameters include the prior constituents and also organic carbon (suspended and dissolved). Standard protocols were used to insure consistency (Shelton, 1994).

Not all water-quality constituents were determined at all sites. The number of samples also varied among the sites (fig. 4) as well as among the constituents. During 1993 and 1994, most sites were sampled between 23 and 25 times, but some sites were sampled as many as 38 times. Water quality sampling at NAWQA sites was usually bimonthly and event based. The numbers of samples per constituent were relatively consistent, but at three non-NAWQA sites only 7-9 samples were analyzed for dissolved organic carbon.

Fish Collection

Fish were collected at each reach (fig. 4) during August and September in 1993 through 1995. The fish community was sampled by electrofishing and seining. Sampling was standardized according to protocols established by Meador and others (1993b) and differed from the reconnaissance sampling. A backpack electrofishing unit (Smith Root Model 12 B) was used in small wadeable streams. The entire reach was electrofished twice (two passes) starting from the downstream boundary. An attempt was made to collect every fish seen. At medium-size wadeable streams two electrofishing passes were made upstream using a stream electrofishing unit (barge type with one or two anodes). Large nonwadeable streams were sampled with an electrofishing boat. Between the two passes, three seine hauls were conducted to sample dominant habitat types and geomorphological units. At some sites no seining was done due to either depth of water or obstacles on the bottom that prohibited effective seining. Seining was not attempted at large nonwadeable sites. Total time of electrofishing was recorded to determine effort so that sampling consistency could be maintained at all sites.

All fish collected were separated by species and counted. All fish were returned to the water except for one or two individuals of each species that were retained as voucher specimens. All voucher specimen field identifications were verified at the University of Minnesota Bell Museum of Natural History where the voucher collection is maintained.

Fish sampling was planned for all reaches all three years; however, only during 1994 were 33 of the total 34 reaches sampled. During the other two years, high water during the late summer precluded sampling at more than 18 reaches. Therefore, only fish community composition data from 1994 are used in this analysis.

Data Analysis

A limitation to the study was the lack of collocated sampling sites. Although every attempt was made to collocate fish sampling sites with stream gaging and water-quality sampling sites, it was not possible to have a complete data matrix for all sites. Some of the sites were selected to be representative of specific land uses within ecoregions or physiographic areas. The sampling design was originally established to compare the fish communities of specific environmental settings and ecoregions within the constraints of stream size. As the project developed, unforeseen factors reduced the completeness of the data matrix. Unusually high discharges all three years precluded reach-level habitat data measurements at three sites. The site selection process identified two sites at which it was necessary to establish streamflow-gaging stations. These two sites became intensive ecological sites with three reaches each. There is no long term hydrologic record for either of these two sites. There was limited availability of soils data for all sites.

Complete data for each of the five data sets (instream habitat, terrestrial habitat, water quality, hydrology, and fish) are not available at every site. Instream habitat data were collected or measured at 19 of the 22 ecological sites where fish were collected or a total of 30 of the 33 reaches where fish were collected. Terrestrial habitat data were available for 14 of the 22 fish collection sites. Soils maps were not available for the other sites. Water-quality data were collected at 11 sites which correspond to fish community sampling sites. Hydrologic variability data were available for 12 sites which corresponded to fish collection sites.

Analysis of data required a stepwise approach that built from single variable correlations to multivariate analyses. Because no hypotheses had been proposed for testing and multivariate analyses are exploratory (Gauch, 1982), the analysis focused on determining correlations among the physical and chemical variables and the fish community. The analyses were used to define general correlations and descriptions of relationships among the physical and chemical characteristics of the environment and fish community composition.

Two multivariate procedures were used on all physical and chemical data sets. The first was cluster analysis (SAS, 1989) which is a classification procedure. Classification procedures assign samples to groups based on the relative distance between them. An average linkage cluster analysis classified the streams by common features and produced groups of sites based on the major discriminating features. Inspection of the results verified that the reaches within cluster groups were similar based on knowledge and experience in the basin. Principal components analysis (PCA), an ordination procedure which separates samples and species on a few orthogonal axes, was then conducted (SAS, 1989) using the same data set as used in the cluster analysis to determine the major environmental variables responsible for the clusters. The loadings (relative amount of variance explained) which resulted from the PCA were then used as input to another cluster analysis. Only variables with relatively high loadings (absolute values >0.20-0.35 depending on the data set) on any of the first three axes were used. The purpose of this analysis was to help identify a single representative environmental variable for each cluster group of variables. Selection of a representative environmental variable from a group of variables identified by correlation, cluster, and PCA was based on which variables had the highest loadings and were judged to have the most biological relevance.

The fish data were also examined with cluster analysis. Jaccard's coefficient of similarity was calculated for each pair-wise combination of reaches sampled. The complements of the coefficients (1-Jaccard coefficient) were used as input for a cluster analysis which identifies the reaches with the most similar species composition. There were two objectives for this analysis. The first was to determine if fish community composition corresponded to either ecoregions or stream size classes. The second was to determine community consistency (at least in terms of species composition). If all three reaches at a site are substantially similar in terms of habitat (instream and terrestrial), then species composition would be expected to be similar if sampling efficiencies are consistent. If fish sampling and habitat were consistent, then all three reaches at a multiple reach sites would be expected to group together in the cluster analysis.

A second ordination procedure was conducted using detrended correspondence analysis (DCA). Unlike the PCA and cluster analysis, which are linear procedures, the DCA is based on a unimodal response of species to environmental gradients (Gauch, 1982; Ter Braak, 1986). The purpose of this analysis was to determine site scores based on the species composition and relative abundance at each site. Although the cluster analysis and PCA indicate which variables of the environmental data sets are most important in distinguishing sites, those variables may not necessarily relate to the fish community. For example, high soil permeability, one of the terrestrial habitat variables, may be quite important in distinguishing sites. However, fish community composition may not relate to high soil permeability. The scores of the sites in species space (results of the DCA first axis, which explains more of the variance in the fish data than the other axes) were correlated to each variable from the environmental data sets to aid in representative variable selection. The correlations should indicate which environmental variables were best correlated with fish species distribution and abundance. The environmental variables which correlated with the DCA scores were compared to the representative variables determined by cluster analysis and PCA. Final selection of representative variables was based on confirmation of importance of a variable by both analyses.

The representative environmental variables from each of the environmental data sets were combined into a single data set, an environmental variable-by-site matrix, and used in a canonical correspondence analysis (CCA) with the fish data. CCA uses the environmental variable-by-site matrix to explain variability in the site-by-species matrix (Ter Braak, 1986). CCA requires fewer variables than samples. Therefore, the unequal number of sampling sites among the four environmental data sets precluded using all the environmental variables in a single analysis with the fish community composition data. To maximize the identification of environmental gradients in the analysis, all variables related to the physical environment (instream and terrestrial habitat, and hydrology) were combined and analyzed with the fish community composition data. Water-quality data (for which there were the fewest sites) were analyzed with the fish data separately. Cluster analysis, principal components analysis, and canonical correspondence analysis have all been recommended for environmental data (Green and Vascotto, 1978).

The model for testing is that major terrestrial landscape factors in the basin, instream habitat features, the chemical environment, and hydrologic variability affect the fish community structure of the streams in the Red River Basin (fig. 5). The model implies that given the context of constant attributes of the basin (soils, topography, mineralogy, basin slope, etc.) and gradually changing physical instream habitat (geomorphological processes as they affect cover, channel width, depth variability, and other factors), the biological community is adapted to predictable variation in environmental factors such as the annual thermal and hydrologic regime. The extremes of climate and hydrologic variability in this basin suggest that the biological communities should be quite resilient to disruption. Reice and others (1990) theorize that a predictable disturbance is not a disruptive factor to a community. Given an environment with a periodic but predictable disturbance, the species in the community of that environment would have adapted to the predictable event (spring floods from snow melt, for example) and not be drastically affected by them. Conversely, unpredictable catastrophic events which occur with irregular frequency (a late summer drought or repetitive floods, for example) may have a significant affect on species composition and abundance. Certain species could be capable of capitalizing on the periodic unpredictable catastrophic event, and have an advantage in this basin where catastrophic hydrologic events occur with unpredictable frequency (Poff and Allen, 1995).

Certainly the model will not be able to account for all the variability in the observed fish communities. Interspecific processes such as competition and predation as well as parasitism and disease occur within communities and have distinct affects on the species composition and relative abundances (Ricklefs, 1987; Power and others, 1988). The amount of variability unexplained by the analysis may be attributed to several of these biotic factors. Unmeasured environmental factors, sampling errors and biases, as well as natural variability will account for much of the unexplained variance in the analysis, but an additional source of disturbance to the biological community results from anthropogenic sources. Some of these have been included in the terrestrial habitat variables (such as land use, the frequency of dams, and the length of ditches in the basin upstream of the sampling site), but others have not been measured and can only be inferred from the analysis. Once the major environmental factors affecting fish community composition have been determined, major sources of the remaining variability may be attributed to biological interactions and anthropogenic factors.

Physical and Chemical Environmental Characteristics

Each of the four environmental data sets were analyzed to reduce the number of variables (table 2) to be used in the CCA with the fish community composition data. Representative variables (table 3) from each of the environmental data sets that distinguished individual sites and reaches were identified through correlation, cluster analysis and principal components analysis. The representative variables were confirmed by correlation to fish species scores (first axis) derived from a DCA. The representative variables had the highest loadings from the PCA (>0.20) and/or correlated ($r>0.4$, $p<0.05$) with the first axis scores from the DCA of fish community composition and relative abundance, or were judged to have the greatest biological significance of any variable within a cluster group if neither of the prior two criteria

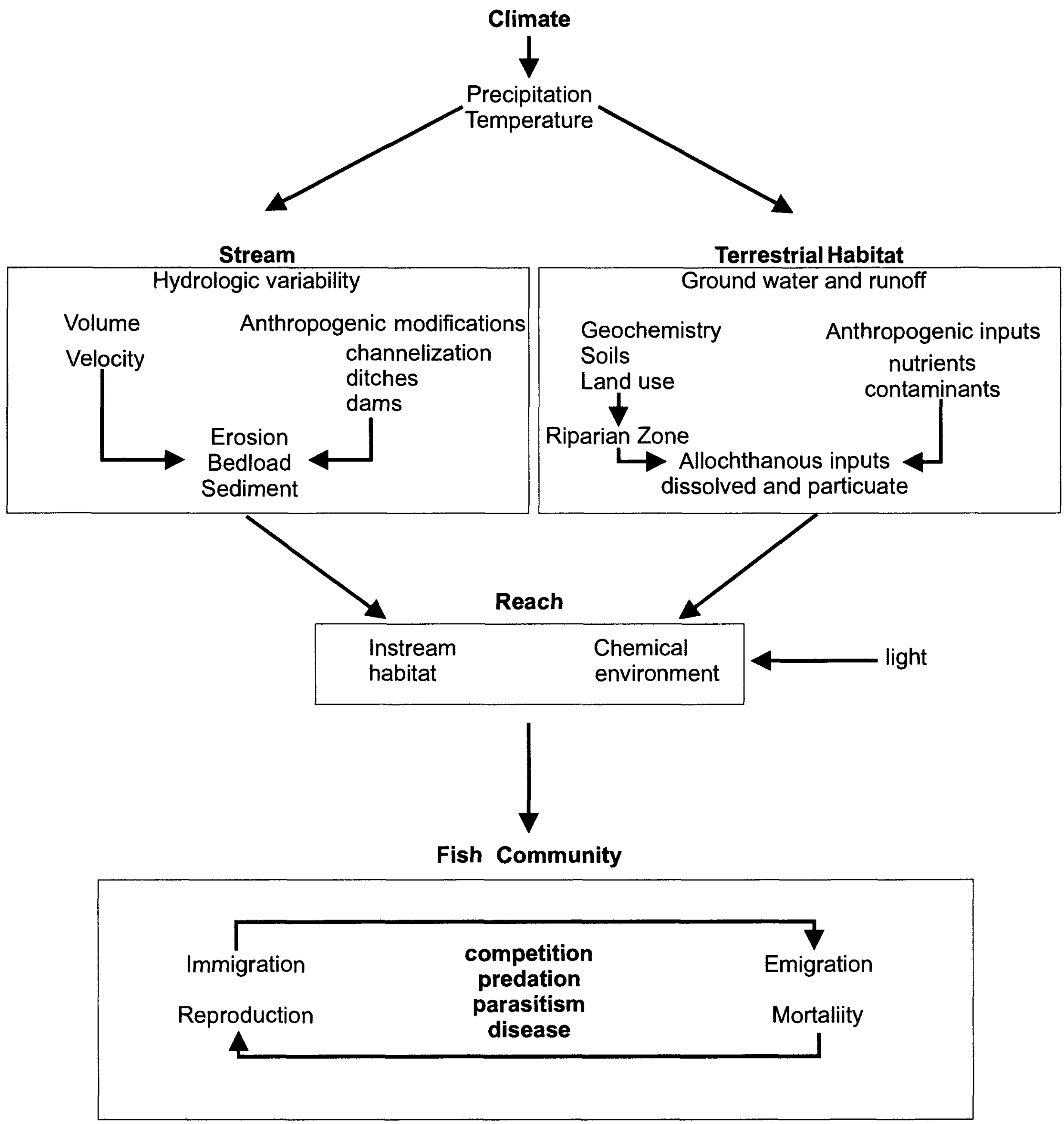

Figure 5. Model of environmental factors affecting fish community composition.

were met. Details of these procedures and analyses that reduced the total number of variables to several representative variables are presented in the Supplemental Information section.

Within the instream habitat data set, mean channel width, number of substrates, coefficient of variation of stream-flow velocity, and erosion frequency were selected as representative variables. Within the terrestrial habitat data set the following were selected as representative variables: left bank riparian zone discontinuity, percent of soils classified as highly erodible, percent as forested land, and percent as riverine wetlands. Hydrologic variability during 1992 through 1994 was represented by the coefficient of variation of April flow and the annual frequency of high flow and low flow events. Representative water-quality variables were percent of suspended sediment less than 2 millimeters in diameter, minimum dissolved oxygen concentration, and maximum concentration of total phosphorus.

Fish Community Composition

Fish were collected from 33 reaches at 22 sites during 1994. Fifty species were found during this study although there are 75 species reported from these rivers in the Red River Basin (Goldstein, 1995). Species richness varied from a low of 7 species in the Snake River to a high of 26 species in Sheyenne River (table 4). Several species were collected only at single locations within one state, but not the other (large scale stoneroller, golden shiner, brassy minnow, brown bullhead, orangespotted sunfish, and bluegill in North Dakota; silver lamprey, blackchin shiner, blacknose shiner, bigmouth buffalo, yellow bullhead, and largemouth bass in Minnesota). Some species were only collected in the mainstem of the Red River (goldeye, mooneye, silver chub, and river shiner). Goldstein (1995) partially attributed differences in species distributions to differences in the number of ecoregions a river flows through and the amount of time since the most recent glaciation in different areas of the Red River Basin. The most common species collected was the blackside darter which was present at 21 of the 22 sites. Blacknose dace, creek chubs, fathead minnows, common shiners, white suckers, and johnny darters were also common species. Each of these species was found at the majority of sites.

There were insufficient sites in each ecoregion to test statistically differences in species composition among ecoregions. When the initial stream size groups were compared, there were more species found at medium-size streams than at small streams ($p<0.016$, sign rank test, Helsel and Hirsh, 1992). No difference was detected between the number of species in large nonwadeable rivers and medium-size streams.

Relation of Physical and Chemical Characteristics to Fish Community Composition

There were insufficient data to allow a combined analysis with all physical and chemical variables. The water-quality data set contained the fewest number of sites and so was analyzed separately. The representative physical variables, including instream habitat, terrestrial habitat, and hydrology were analyzed with the fish community data by CCA. None of the mainstem sites or large nonwadeable sites (except the three Roseau River reaches) were included in this analysis due to lack of terrestrial habitat data, primarily soils data. The large river sites were shown to have a different species composition by the DCA (Supplemental Information section). The gradient of species associations with river size could possibly have confounded the other physical gradients in the data.

Prior to running the CCA, a correlation matrix was constructed for the representative environmental variables. The mean annual seven day high flow was significantly correlated to mean channel width ($r=0.8$, $p<0.001$) indicating that as stream size increases the high discharge values also increase. Therefore, to reduce redundancy, the mean annual seven day high flow variable was deleted from further analysis.

CCA axes are constrained as linear combinations of the environmental variables. Therefore, the fish species are directly related to a group of environmental variables which comprise the axes. The physical characteristics are the environmental variables in this analysis. The location of a species relative to an axis is at the species' optimum set of conditions that make up the axis. That location is the mode of the unimodal distribution of that species abundance in the gradient expressed by the axis.

The CCA of physical characteristics of streams and fish community composition was able to explain much of the variance in the data sets. The total amount of variance explained in the species-environment relation on four axes was 76 percent. Most of the variance was explained by Axes 1 and 2 (table 5) which accounted for 27.7 and 24.8 percent of the variance, respectively. Axes 3 and 4 accounted for 15.2 and 8.4 percent, respectively. The four axes combined explained about 63 percent of the variance in the species data (table 5). The eigenvalues for Axes 1 and 2 were similar (0.383 and

Table 2.—Physical and chemical variables determined for sampling sites in the Red River of the North study unit.

Data set	Variable	Variable abbreviation	Units or scores
Instream habitat	Mean channel width	I-mchwidth	meters
	Mean stream flow velocity	I-mvelocity	centimeters per second
	Mean depth	I-mdepth	centimeters
	Mean canopy angle	I-mcanangle	degrees
	Coefficient of variation of channel width	I-cvchwidth	
	Coefficient of variation of velocity	I-cvvelocity	
	Coefficient of variation of depth	I-cvdepth	
	Coefficient of variation of canopy angle	I-cvcanangle	
	Mean width to depth ratio	I-mw/d	
	Number of different substrate types	I-nsubstr	count
	Frequency of linear bank shape of left and right banks	I-lbshp and I-rbshp	number of observations of linear shape
	Sum of bank stability of left and right banks	I-lbstab and I-rbstab	scale of 1 to 4; 4 is most stable
	Frequency of erosion of left and right banks	I-lberosfq and I-rberosfq	number of observations of erosion /total observations
	Frequency of bank erosion, left and right banks combined	I-erosfq	number of observation for both banks/ total observations
	Percent area with instream cover	I-instrcov	percent of total area
	Reach gradient	I-rslope	centimeters per kilometer
	Reach sinuosity	I-sinu	
Terrestrial habitat	Soil characteristics		
	shallow soil permiability (low, moderate, and high)	T-lperm, T-mperm, T-hperm	percent of total area; three classes: low, moderate and high
	soil erosion factor (low, moderate, and high)	T-Kl, T-Km, T-Kh	percent of total area; three classes: low, moderate and high
	soil wind erodibility (low, moderate, and high)	T-lwinderos, T-mwinderos, T-hwinderos	percent of total area; three classes: low, moderate and high
	Land use/cover in groups		
	urban	T-L1	percent of total area
	agricultural	T-L2	percent of total area
	forest	T-L4	percent of total area
	water	T-L5	percent of total area
	wetland	T-L6	percent of total area
	lacustrine wetlands	T-WLwet	percent of total area
	riverine wetlands	T-WRwet	percent of total area

Table 2.—Physical and chemical variables determined for sampling sites in the Red River of the North study unit—Continued.

Data set	Variable	Variable abbreviation	Units or scores
	palustrine wetlands	T-WPwet	percent of total area
	other		
	Mean width of riparian zone (left and right banks)	T-lbripwidth T-rbripwidth	meters
	Coefficient of variation of riparian zone width (left and right banks)	T-lbcvripwidth T-rbcvripwidth	
	Tree density of riparian zone (left and right banks)	T-lbtreeden T-rbtreeden	trees/area
	Bank riparian discontinuity (left and right banks)	T-lbbrks T-rbbrks	percent of total bank length without riparian trees
	Stream gradient	T-sslope	centimeters per kilometer
	Number of downstream dams	T-dsdams	number
	Number of tributaries	T-tribs	number
	Length of ditches	T-ditches	kilometers
Hydrology	Monthly mean streamflow	H-qm(month)	cubic meters per second
	Coefficient of variation of monthly mean streamflow	H-qcv(month)	
	Frequency of bankfull stage (two year recurrence)	H-qbf	
	Mean annual frequency of high streamflow (greater than the 75th percentile of period of record)	H-q75	
	Mean duration (days) of high streamflow (greater than the 75th percentile of period of record)	H-dq75	days
	Mean annual frequency of low streamflow (less than the 25th percentile of period of record)	H-q25	
	Mean duration (days) of low streamflow (less than the 25th percentile of period of record)	H-dq25	days
	Mean number of days of 0 streamflow	H-daq0	number
	Mean Julian date streamflow (high and low)	H-hmjulq H-lmjulq	
	One day high streamflow	H-qh1	cubic meters per second
	Three day high streamflow	H-qh3	cubic meters per second
	Seven day high streamflow	H-qh7	cubic meters per second
	Thirty day high streamflow	H-qh30	cubic meters per second
	Ninety day high streamflow	H-qh90	cubic meters per second
	One day low streamflow	H-ql1	cubic meters per second
	Three day low streamflow	H-ql3	cubic meters per second
	Seven day low streamflow	H-ql7	cubic meters per second
	Thirty day low streamflow	H-ql30	cubic meters per second

Table 2.—Physical and chemical variables determined for sampling sites in the Red River of the North study unit—Continued.

Data set	Variable	Variable abbreviation	Units or scores
	Ninety day low streamflow	H-ql90	cubic meters per second
	Mean high streamflow pulse count	H-qhpc	cubic meters per second
	Mean low streamflow pulse count	H-qlpc	cubic meters per second
Water quality	Specific conductance (maximum and minimum)	W-sp (mx or mn)	microseimens at 25 degrees Celsius
	pH (maximum and minimum)	W-pH (mx or mn)	
	Temperature (maximum and minimum)	W-temp (mx or mn)	degrees Celsius
	Dissolved oxygen (maximum and minimum)	W-do (mx or mn)	milligrams per liter
	Dissolved calcium (maximum and minimum)	W-calc (mx or mn)	milligrams per liter
	Dissolved sulfate (maximum and minimum)	W-sulf (mx or mn)	milligrams per liter
	Dissolved chloride (maximum and minimum)	W-chlor (mx or mn)	milligrams per liter
	Dissolved nitrate plus nitrite (maximum and minimum)	W-nitr (mx or mn)	milligrams per liter
	Total ammonia and organic nitrogen (maximum and minimum)	W-amm+org (mx or mn)	milligrams per liter
	Total inorganic nitrogen (maximum and minimum)	W-innitr (mx or mn)	milligrams per liter
	Total phosporus (maximum and minimum)	W-p (mx or mn)	milligrams per liter
	Total suspended sediment (maximum and minimum)	W-sed (mx or mn)	milligrams per liter
	Fine particulates (maximum and minimum)	W-%fines (mx or mn)	percent smaller than sand (less than 2 millimeters)
	Dissolved organic carbon (maximum and minimum)	W-doc (mx or mn)	milligrams per liter

0.342, respectively) and indicated that both axes were important in the ordination. The weighted correlation matrix of the physical characteristics and the ordination axes (table 6) indicated that mean channel width, coefficient of variation of velocity, number of substrates, riverine wetlands, and coefficient of variation of mean April flow were important on the first axis. Mean channel width and the mean annual frequency of low flow were important on the second axis. Mean channel width had a moderately high inflation factor (17.9) which means it was correlated to other variables. The weighted correlation matrix (physical characteristic by physical characteristic) indicated that mean channel width was correlated to the number of substrates (r=0.80). Both variables were retained in the analysis because of their biological significance to the fish community and understanding of the ordination. The Monte Carlo permutation test indicated that the first axis was not significant (p=0.15). This was due to the second axis having an eigenvalue similar to the first axis such that the ratio of the variance explained by the first axis to the total variance (inertia) would not be significant (approaching 1). The overall test of the CCA was significant (p=0.01).

The physical characteristics may be ranked in importance to the ordination and the significance of

Table 3.—Representative variables from each environmental data set and the variables represented.
[CV, coefficient of variation]

Data set	Representative variables			
Instream habitat	**Mean channel width**	**Number of substrates**	**Coefficient of variation of velocity**	**Erosion frequency**
Variables represented	mean depth	instream cover	CV depth	reach gradient
	mean streamflow velocity	bank stability	CV canopy angle	left bank shape
	mean canopy angle		right bank shape	
Terrestrial habitat	**Left bank riparian discontinuity**	**Percent area with high soil erosion factor**	**Percent area as forested land**	**Percent area as riverine wetlands**
Variables represented	number of downstream dams	percent area with low soil permeability	percent area with moderate soil erodibility	right bank riparian tree density
	stream gradient	number of tributaries	percent area with high soil permeability	percent area as lacustrine wetlans
	percent area with soils with low wind erodibility	percent area with moderate soil wind erodibility	left bank mean riparian zone width	
	CV left bank riparian zone width	right bank riparian discontinuity	percent area as palustrine wetlands	
		CV right bank riparian zone width		
Hydrology	**CV April flow**	**Frequency of high flows**	**Freqency of low flows**	
Variables represented	CV mean monthly flow			
	high flows for various time periods			
	low flows for various time periods			
	mean julian dates of high and low flows			
Water quality	**Minimum percent of suspended sediment smaller than sand**	**Minimum dissolved oxygen conentration**	**Maximum total phosphorus concentration**	
Variables represented	minimum dissolved organic carbon	minimum dissolved calcium	maximum and minimum dissolved sulfate	maximum dissoved calcium
	minimum nitrogen as ammonia and organic		maximum and minimum dissolved chloride	minimum total phosphorus

Table 4.—Fish species collected in the Red River of the North and its major tributaries.

[Site number corresponds to site desciptions on Table 1 and locations on Figure 4]

Fish species by Family		Minnesota												Mainstem			North Dakota						
Common Name	Scientific name	Buffalo River (19)	Clearwater River at Red Lake Falls (13)	Clearwater River near Berner (20)	Ottertail River (21)	Red Lake River at Crookston (8)	Red Lake River at Red Lake Falls (9)	Roseau River (3)	Sandhill River (22)	Snake River (5)	Two Rivers (16)	Wild Rice River at Mahnomen (18)	Wild Rice River at Twin Valley (4)	Red River at Joliette (1)	Red River below Fargo (7)	Bios de Sioux River (11)	Forest River (17)	Goose River (14)	Maple River (15)	Pembina River (10)	Sheyenne River at Lisbon (2)	Sheyenne River at Warwick (12)	Turtle River (6)
Lampreys	***Petromyzontidae***																						
Chestnut Lamprey	*Ichthyomyzon castaneus*						X	X				X	X					X					
Silver Lamprey	*Ichthyomyzon unicuspis*						X																
Mooneyes	***Hiodontidae***																						
Goldeye	*Hiodon alosoides*													X	X								
Mooneye	*Hiodon tergisus*													X	X								
Minnows	***Cyprinidae***																						
Carp	*Cyprinus carpio*	X				X	X	X						X	X	X		X	X		X		
Largescale Stoneroller	*Campostoma oligolepis*																X						
Longnose Dace	*Rhinichthys cataractae*	X	X	X			X		X			X	X				X			X	X	X	X
Blacknose Dace	*Rhinichthys atratulus*		X	X	X		X		X		X	X	X				X		X	X	X	X	X
Creek Chub	*Semotilus atromaculatus*	X	X	X		X	X		X	X	X	X	X				X	X	X	X	X	X	X
Silver Chub	*Macrhybopsis storeriana*													X	X								
Hornyhead Chub	*Nocomis biguttatus*	X	X		X	X	X					X	X				X						
Northern Redbelly Dace	*Phoxinus eos*			X					X			X	X										
Golden Shiner	*Notemigonus crysoleucas*																					X	
Fathead Minnow	*Pimephales promelas*	X	X	X		X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Bluntnose Minnow	*Pimephales notatus*						X										X				X		
Brassy Minnow	*Hybognathus hankinsoni*																			X			
Spottail Shiner	*Notropis hudsonius*				X	X	X							X							X	X	
Common Shiner	*Luxilus cornutus*	X	X	X	X	X	X	X	X		X	X	X				X	X	X	X		X	X
Spotfin Shiner	*Cyprinella spiloptera*	X												X	X	X		X			X		
Emerald Shiner	*Notropis atherinoides*					X	X							X	X	X							
Bigmouth Shiner	*Notropis dorsalis*	X							X			X	X				X	X	X	X			X
River Shiner	*Notropis blennius*													X									
Rosyface Shiner	*Notropis rubellus*	X																			X		
Blackchin Shiner	*Notropis heterodon*				X																		
Blacknose Shiner	*Notropis heterolepis*			X																			
Sand Shiner	*Notropis stramineus*	X				X	X					X	X	X	X	X		X		X	X	X	X
Suckers	***Catostomidae***																						
Bigmouth Buffalo	*Ictiobus cyprinellus*					X																	
Quillback	*Carpiodes cyprinus*													X	X			X					
Silver Redhorse	*Moxostoma anisurum*	X					X								X	X		X			X		
Shorthead Redhorse	*Moxostoma macrolepidotum*	X	X		X	X	X				X		X	X	X			X		X	X		
Golden Redhorse	*Moxostoma erythrurum*	X	X				X	X				X	X								X		
White Sucker	*Catostomus commersoni*	X	X	X	X	X	X	X	X	X	X	X	X			X	X	X	X	X	X	X	X
Catfish	***Ictaluridae***																						
Channel Catfish	*Ictalurus punctatus*													X				X			X		
Brown Bullhead	*Ameiurus nebulosus*																					X	

Table 4.—Fish species collected in the Red River of the North and its major tributaries—Continued.

Fish species by Family		Minnesota												Mainstem			North Dakota						
Common Name	Scientific name	Buffalo River (19)	Clearwater River at Red Lake Falls (13)	Clearwater River near Berner (20)	Ottertail River (21)	Red Lake River at Crookston (8)	Red Lake River at Red Lake Falls (9)	Roseau River (3)	Sandhill River (22)	Snake River (5)	Two Rivers (16)	Wild Rice River at Mahnomen (18)	Wild Rice River at Twin Valley (4)	Red River at Joliette (1)	Red River below Fargo (7)	Bios de Sioux River (11)	Forest River (17)	Goose River (14)	Maple River (15)	Pembina River (10)	Sheyenne River at Lisbon (2)	Sheyenne River at Warwick (12)	Turtle River (6)
Black Bullhead	*Ameiurus melas*	X			X											X	X	X	X			X	X
Yellow Bullhead	*Ameiurus natalis*				X																		
Stonecat	*Noturus flavus*	X																		X			
Tadpole Madtom	*Noturus gyrinus*																X		X	X	X	X	X
Pikes	***Esocidae***																						
Northern Pike	*Esox lucius*	X			X	X		X						X			X	X		X	X	X	
Mudminnows	***Umbridae***																						
Central Mudminnow	*Umbra limi*			X		X		X	X	X		X	X										
Trout-Perch	***Percopsidae***																						
Trout-Perch	*Percopsis omiscomaycus*					X		X				X	X	X	X					X			
Cods	***Gadidae***																						
Burbot	*Lota lota*							X						X									
Sticklebacks	***Gasterosteidae***																						
Brook Stickleback	*Culaea inconstans*			X					X	X		X						X		X	X	X	X
Drums	***Scianidae***																						
Freshwater Drum	*Aplodinotus grunniens*							X						X	X	X		X					
Bass and Sunfishes	***Centrarchidae***																						
Largemouth Bass	*Micropterus salmoides*				X																		
Smallmouth Bass	*Micropterus dolomieu*		X			X	X														X		
Orangespotted Sunfish	*Lepomis humilis*																				X		
Bluegill	*Lepomis macrochirus*																				X		
Rock Bass	*Ambloplites rupestris*	X	X			X	X				X	X	X	X	X		X			X	X		X
Black Crappie	*Pomoxis nigromaculatus*				X			X								X					X		
Perches	***Percidae***																						
Yellow Perch	*Perca flavescens*				X	X	X	X								X		X					X
Walleye	*Stizostedion vitreum*						X	X						X	X			X			X		
Sauger	*Stizostedion canadense*							X						X	X								
Blackside Darter	*Percina maculata*	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X
Johnny Darter	*Etheostoma nigrum*	X		X	X	X	X	X	X	X	X	X	X	X		X	X			X	X	X	X
Iowa Darter	*Etheostoma exile*			X									X			X			X			X	
Total species		20	12	13	15	19	22	17	12	7	9	18	19	22	16	14	16	20	11	18	26	17	15

each variable tested. A constrained CCA relates one environmental variable to the first axis. When all other environmental variables are omitted from the analysis, the ratio of the eigenvalue of the first axis to the second axis (unconstrained) is an indication of the amount of variance explained by that one variable. A Monte Carlo permutation test then tests the significance of the ratio. The physical characteristics with high ratios of eigenvalues from Axis 1 to Axis 2 were percent of land as riverine wetlands, mean channel width, coefficient of variation of April flow, number of substrates, and mean annual frequency of high flow (table 7). All of these variables were significant in the ordination ($p \leq 0.03$). The environmental variables can also be tested in a partially constrained CCA. In this test, one environmental variable is related to the first axis and all the other variables are made covariables. The same ratio of eigenvalues is determined and tested by the Monte Carlo permutation test. The purpose of this test was to determine the effect of a single environmental variable

Table 5.—Summary of canonical correspondence analysis (CCA) of representative physical variables and fish community composition.

Summary statistic	Axis 1	Axis 2	Axis 3	Axis 4
Eigenvalues	0.383	0.342	0.210	0.117
Species and environment correlations	.956	.968	.983	.949
Cumulative percent of variance explained in species data	23.3	43.3	56.1	63.3
Cumulative percent of variance explained in species-environment relation	27.7	52.5	67.7	76.1

Table 6.—Weighted correlation matrix of representative physical variables and canonical correspondence analysis (CCA) axes.

Representative physical variable	Correlation coefficient			
	Axis 1	Axis 2	Axis 3	Axis 4
Coefficient of variation of velocity	0.4059	-0.1228	0.0768	-0.2927
Soil erosion factor-high	.2904	-.2459	.3158	-.0787
Landcover of forest	-.1830	.3001	-.1267	-.0028
Bank riparian discontinuity-left bank	.0339	-.1150	.1318	-.3845
Mean channel width	.4744	.4788	-.2396	-.2545
Number of substrates	.4980	.2721	-.4450	-.3504
Frequency of erosion-left and right bank combined	.0655	.2074	.1515	.7236
Riverine wetlands	.6926	.1331	-.2028	-.0239
Coefficient of variation of mean April flow	.6243	-.0384	.1520	-.1215
Mean annual frequency of flow greater than the 75th percentile of period of record	-.2542	.2389	.8299	-.1884
Mean annual frequency of flow less than the 25th percentile of period of record	.0445	-.4631	.1330	-.2962

relative to the combined effect of the other environmental variables. In this analysis (table 8), the mean annual frequency of high flow, the frequency of erosion, and mean channel width were the most significant variables (eigenvalue ratios >1, p=0.01), although the coefficient of variation of April stream-flow and the percent of land as riverine wetlands were also significant (eigenvalue ratio >1, p=0.03). The physical characteristic with the least effect on the ordination appears to be the number of substrates, however this more likely due to the high correlation between the number of substrates and mean channel width.

The graphic representation of the CCA analysis is a biplot. The biplot contains two sets of data, one is the axes determined by the environmental variables and the other is the ordination of either fish species composition or sampling sites relative to those axes. The environmental variable arrows indicate the relation of an environmental variable to an axis by the angle between the arrow and an axis and the importance of the variable in the ordination by the length of the arrow. The arrow points in the direction of the maximum rate of change of the variable. The biplot (fig. 6) of the physical characteristics and fish species indicate most of the physical characteristics are related to Axis 1. Arrows for

Table 7.—Test for significance of constrained physical variables.
[p is significance level; *, significant p values (less than 0.05)].

Physical variable	Eigenvalue axis 1/ Eigenvalue axis 2	p
Coefficient of variation of velocity	0.32	0.08
Soil erosion factor-high	.33	.11
Landcover of forest	.22	.41
Bank riparian discontinuity-left bank	.17	.59
Mean channel width	.62	.01*
Number of substrates	.54	.02*
Frequency of erosion-left and right bank combined	.24	.29
Riverine wetlands	.63	.01*
Coefficient of variation of mean April flow	.54	.01*
Mean annual frequency of flow greater than the 75th percentile of period of record	.48	.03*
Mean annual frequency of flow less than the 25th percentile of period of record	.29	.16

Table 8.—Partial canonical correspondence analysis (CCA).
[p is significance level; *, significant p values (<0.05)].

Physical variable	Eigenvalue axis 1/ Eigenvalue axis 2	p
Coefficient of variation of velocity	0.67	0.23
Soil erosion factor-high	.64	.16
Landcover of forest	.97	.03*
Bank riparian discontinuity-left bank	.74	.17
Mean channel width	1.32	.01*
Number of substrates	.34	.63
Frequency of erosion-left and right bank combined	1.82	.01*
Riverine wetlands	1.19	.03*
Coefficient of variation of mean April flow	1.01	.03*
Mean annual frequency of high flow (greater than the 75th percentile of period of record)	1.96	.01*
Mean annual frequency of low flow (less than the 25th percentile of period of record)	.99	.03*

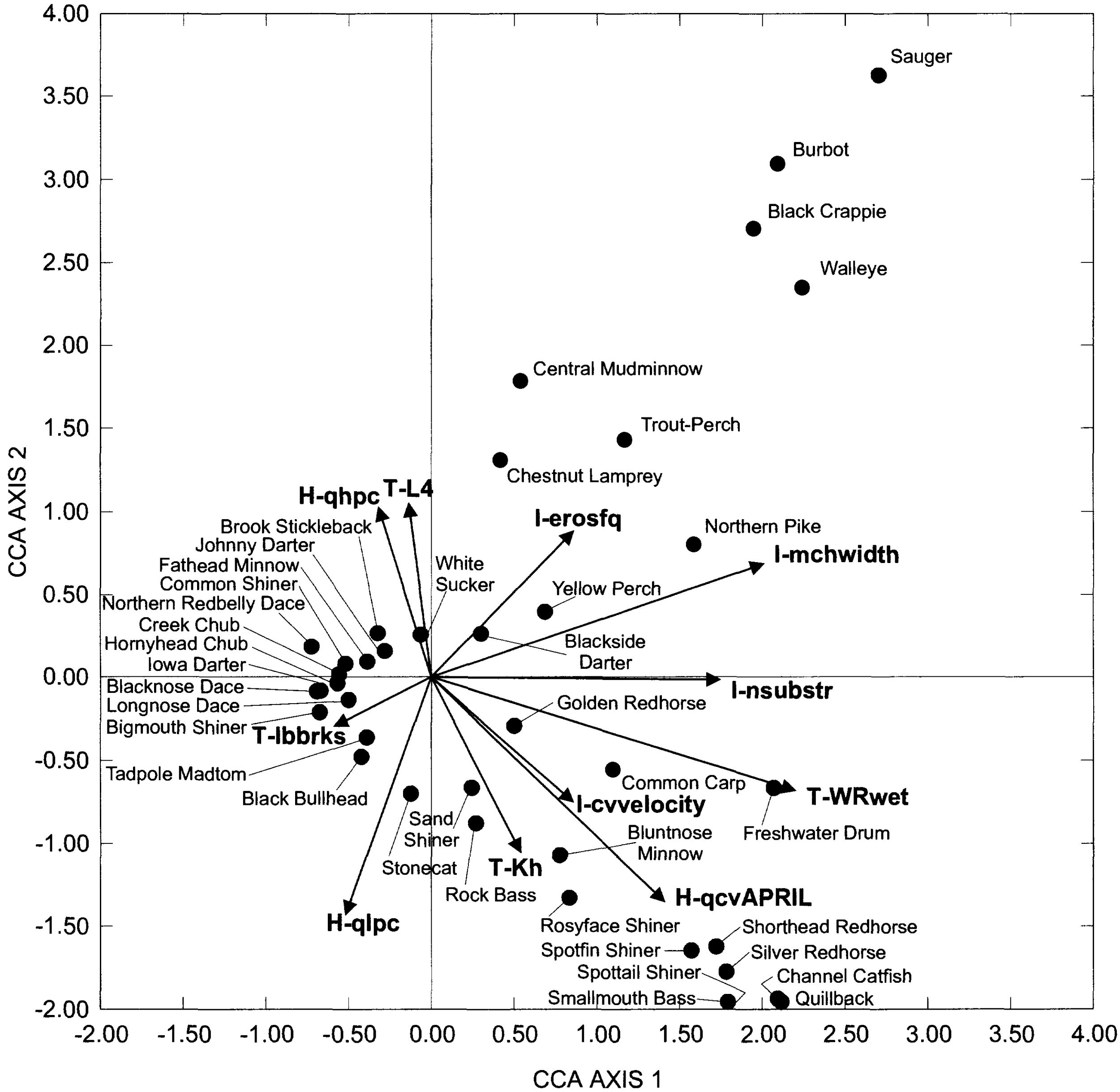

Figure 6. Canonical correspondence analysis (CCA) biplots of representative environmental variables in relation to fish community composition. [Refer to table 2 for definitions of variable abbreviations.]

most of the physical characteristics point in a similar direction along Axis 1. Although the number of substrates (I-nsubstr) is most closely aligned with the first axis, mean channel width (I-mchwidth), variability in spring (April) stream-flow (H-qcvAPRIL), variability of streamflow velocity (I-cvvelocity), and percent of land cover as riverine wetlands (T-WRwet) all have a similar influence on the ordination of fish species based on the relative length of their environmental arrows. Erosion frequency (I-erosfq) and the relative amount of highly erodible soils (T-Kh) appear to align between Axes 1 and 2, but based on the length of the corresponding arrows, these physical factors are not as influential in the ordination as the other physical characteristics. The frequency of high (H-qhpc) and low flows (H-qlpc) and the percent of forested land (T-L4) are inversely related to axis 1 variables. In small streams, changes in the amount of flow would be more rapid than in larger rivers. Similarly, riparian zone discontinuity (T-lbbrks) would be more prevalent at small stream sites in this highly agricultural basin. Nonagricultural land cover which is not classified as wetlands would be forested, particularly near the small streams.

All the physical characteristics used in the CCA are representative environmental variables. Mean channel width also represents mean depth, mean streamflow velocity, and mean canopy angle (table 3). The number of substrates represents bank stability (hence the inverse relation to erosion frequency), and instream cover. The instream habitat variables produce a gradient of stream size and habitat complexity that includes overhanging bank vegetation, substrate particle size diversity, and instream structures such as woody debris.

The first axis of the ordination is strongly correlated to the percent of riverine wetlands (T-WRwet) upstream of the reach sampled for fish. The variable also represents lacustrine wetlands and riparian zone tree density. The relative amount of forested land (T-L4) represents soils with high permeability and moderate erodibility and the width of the riparian zone (table 3). Within the 4,000 m distance class many of the small streams had wide riparian zones and so were also classed as having a greater amount of forested land. These terrestrial habitat characteristics tend to ordinate sites by the amount of nonagricultural land and then distinguish sites further by whether that land is forested or wetlands.

The coefficient of variation of mean monthly flow in April (H-qcvAPRIL) is representative of the variation in discharge for the other months of the year (table 3). The frequency of high (H-qhpc) and low (H-qlpc) flows were used to indicate the atypical hydrologic regime experienced in the Red River Basin during the three years prior to sampling. The normal precipitation gradient in the basin, which increases from west to east and influences the frequency of high and low flows geographically, was evident during this period. The ordination of the hydrologic variables indicated that the variability in flow during the spring (when fish spawning migrations and reproduction occur) influences the distribution of fish more than the frequency of high or low flow events.

The species distributions reflect several gradients and identify preferential conditions for Red River Basin fish species along those gradients. The first is stream size. Species to the right of the origin of the axes are found in larger streams while species to the left of the origin are found in smaller streams (fig. 6). Sauger, burbot, walleye, trout-perch, black crappie, and northern pike are more commonly found (are closer to their environmental optima in the Red River Basin) in medium size and larger rivers (mean channel width, I-mchwidth) with diverse substrates (number of substrates, I-nsubstr), eroded banks (erosion frequency, I-erosfq), and more nonagricultural land use (percent forested land, T-L4; and percent wetlands, T-WRwet). The flow of these rivers is not as flashy in the spring but they tend to have a higher frequency of high discharges (high flow pulse count, H-qhpc) and a lower frequency of low flows (low flow pulse count, H-qlpc)). Species such as quillback, freshwater drum, channel catfish, smallmouth bass, common carp, spottail and spotfin shiners, and silver and shorthead redhorse are also fish of larger medium-size rivers. They tend to be more commonly found in rivers that have more variable spring discharge (coefficient of variation of April flow, H-qcvAPRIL) and are subject to more frequent periods of low flow. Most of the remaining fish are smaller stream species. These include blacknose and longnose dace, creek chubs, fathead minnow, Iowa and Johnny darters, bigmouth shiner, common shiner, and hornyhead chub. These species are more frequently found in streams in a more agricultural landscape with less variable substrates. These species tend to prefer streams with less variable spring discharge and more frequent low-flow periods.

The ordination of the sampling sites and environmental variables (fig. 7) also indicates the stream-size gradient. The larger medium-size wadeable streams are located to the right of the origin of the axes and the smaller streams to the left. All the reaches at a multiple reach site grouped together. The Roseau River in the Northern Minnesota Wetlands ecoregion was most different from the other sites and locations. The Roseau River reaches were among the widest in this group of sites (the mainstem Red River and Red Lake River at Crookston were large, nonwadeable sites but not included in this analysis). The Roseau River site also exhibited the greatest diversity of substrate sizes ranging from silt to boulders.

The water-quality data were also analyzed with CCA and the fish community composition data. Although there were fewer sites in the data set, the range of stream sizes was greater. The large mainstem Red River sites were included in this analysis.

The representative water-quality variables on the first two CCA axes accounted for nearly 87 percent of the variability in the fish community composition-environment relation data which explains almost 30 percent of the variance in the fish data (table 9). Minimum values of the percent of suspended sediment smaller than sand (fine particulates, W-%Finesmn) and minimum dissolved oxygen concentration (W-domn) correlated with axis 1 and maximum total phosphorus concentration (W-pmx) correlated with axis 2 (fig. 8). The eigenvalues of the axes were 0.435 and 0.214, respectively, which indicates that Axis 1 was most

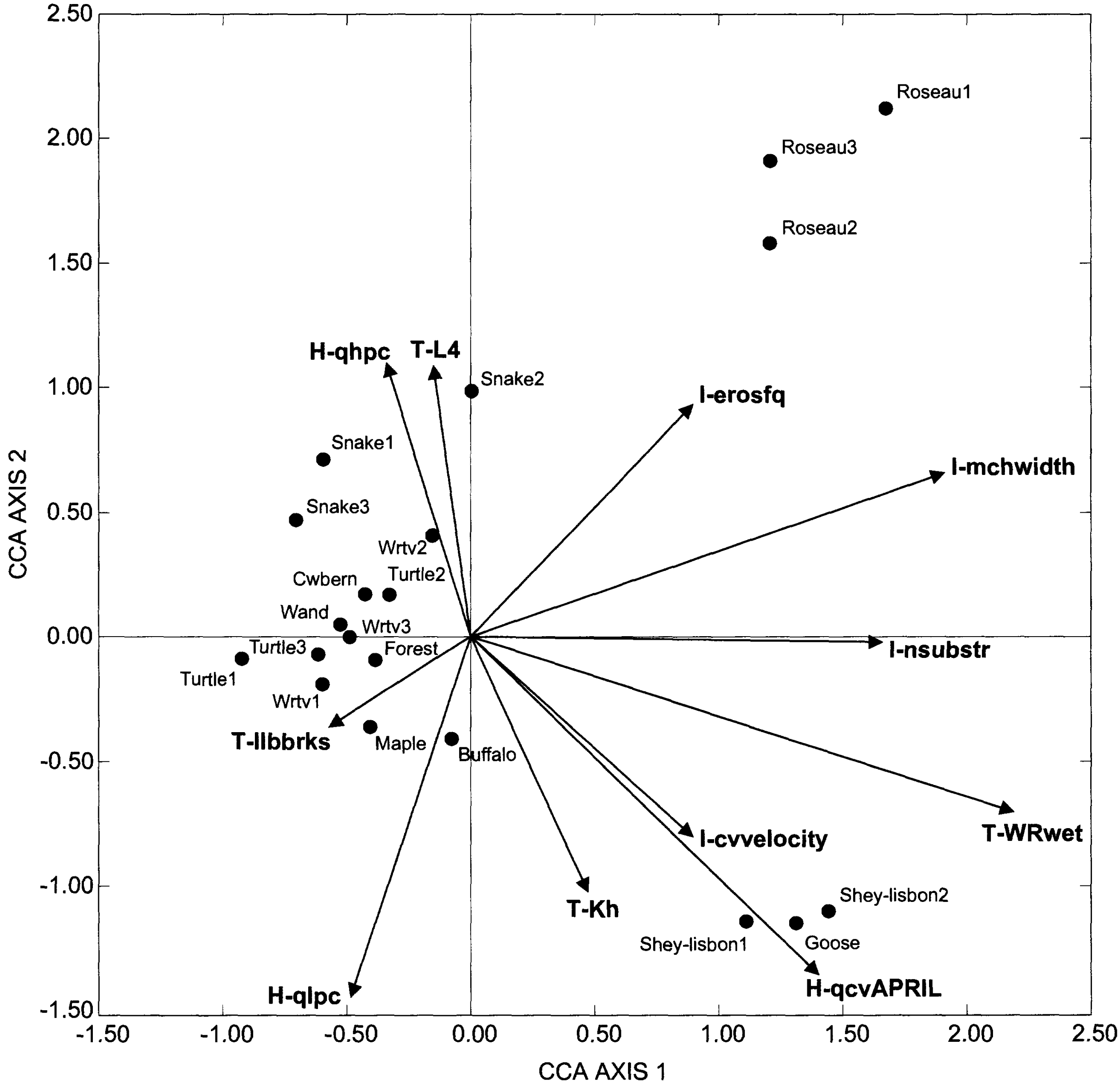

Figure 7. Canonical correspondence analysis (CCA) biplots of representative environmental variables in relation to sampling sites. [Refer to table 2 for definitions of variable abbreviations and table 1 for definitions of site abbreviations.]

influential in the ordination. The interset correlations of environmental variables with Axis 1 indicated that the minimum proportion of fine material was highly correlated with the first axis (r=0.91) as was minimum dissolved oxygen (r=-0.77). The minimum proportion of fine suspended sediment and minimum dissolved oxygen were significant on the first axis (t statistics >2.1, p=0.05) and all three representative variables were significant on the second axis. Although a t-test is not strictly appropriate in indirect (unimodal) statistics, it is indicative of importance. The Monte Carlo permutation test indicated the first axis was significant (p=0.01).

The minimum percent of suspended sediment smaller than sand was a representative variable for minimum nitrogen (ammonia plus organic) and minimum dissolved organic carbon (table 3). Minimum dissolved oxygen was representative of minimum dissolved calcium. Maximum total phosphorus represented minimum total phosphorus, maximum and minimum concentrations of sulfate and chloride and maximum

Table 9. Summary of canonical correspondence analysis (CCA) of water-quality and fish community composition.

Summary statistic	Axis 1	Axis 2	Axis 3
Eigenvalues	0.435	0.214	0.098
Species and environment correlations	.936	.925	.869
Cumulative percent of variance explained in species data	19.9	29.8	34.3
Cumulative percent of variance explained in species-environment relation	58.2	86.9	100.0

calcium, and maximum dissolved nitrogen (ammonia plus organic). The first axis (fig. 8) appears to reflect a gradient of turbidity, the relative proportion of fine particulate material and a range of low dissolved oxygen concentrations that may stem from the prolonged winter with snow and ice cover. The second axis reflects the wide range of chemical and nutrient concentrations as evidenced by the ordination of both maximum and minimum dissolved ion concentrations. In a highly agricultural basin this would not be unusual. The high concentrations of nutrients are associated with high flows in the spring after the fields have been fertilized (Tornes and Brigham, 1994).

Implications for Water-Quality Assessment

The status of water quality in the Red River Basin as indicated by fish community composition and abundance appears to be adequate to support a diverse community. However, physical and chemical characteristics of the basin affect the composition of fish communities (fig. 9). Low concentrations of dissolved oxygen and fine suspended sediment have a negative effect on some fish species affecting both respiration (gill occlusion) or reproduction (smothering of eggs). These two conditions generally do not occur simultaneously. Most low oxygen concentrations occur during winter under ice or during periods of very low to zero flow, while high concentrations of fine sediment normally occur during periods of high flow. Increased erosion that contributes to fine sediment and biological oxygen demand may be mediated during summer by wetlands which act to filter sediment and, through primary production, add dissolved oxygen. A continuous, densely vegetated, riparian zone reduces sediment input from highly erodible soils and provides canopy cover and instream cover from root wads, woody debris, and undercut banks as well as providing organic matter from leaf fall. The small streams in the Red River Basin are subjected to greater hydrologic variability than the larger rivers; hydrologic variability becomes less important as river size increases and acts as a buffer to variable conditions.

Investigation of the effect of biological interactions on community composition was beyond the scope of this report. Although competition and predation can have significant effects on fish community composition and habitat selection (Menge and Sutherland, 1976; Werner and others, 1977, Schlosser, 1987; Power and others, 1988), the emphasis was on environmental or abiotic factors that affect fish community composition in streams. The results of the analysis indicate that in streams of the Red River Basin, abiotic factors have a significant role and may be expressed in terms of species composition and abundances.

Instream habitat features have been shown to correlate with fish community composition and species abundances. Several features of physical habitat that have been correlated with fish community composition and abundance are depth, substrate types, current, and canopy cover (Sheldon, 1968; Horwitz, 1978; Gorman and Karr, 1978; Murphy and others, 1982; Schlosser, 1982). All these factors have been referred to in the context of habitat diversity. In the Red River Basin, these same factors were important in site distinction and correlated to fish community composition. The results of the cluster analysis and PCA (Supplemental Information section) indicated that the most useful measures for site distinction were those factors that relate to habitat volume (stream size) and instream habitat diversity. The correlations of the DCA fish community scores with reach-level habitat variables also indicated size related factors of depth and channel width as most correlated with the fish community ordination (Supplemental Information section).

Instream habitat measures that relate to stream geomorphology (bank stability and bank shape) were only moderately useful for site distinction. The low gradient of the Red River Basin may preclude these factors from being useful for site discrimination. In a comparison of ecoregions by both physical habitat

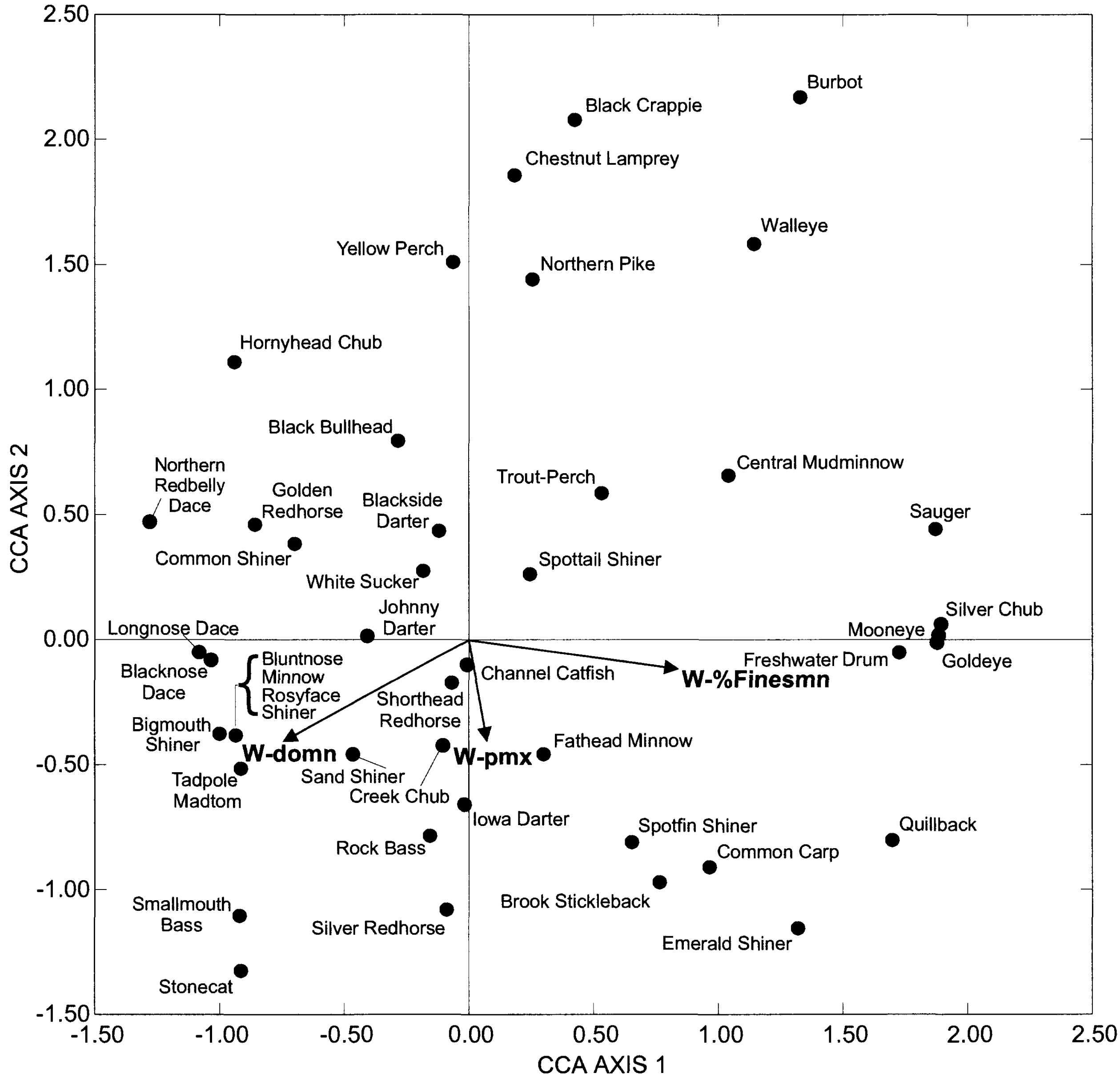

Figure 8. Canonical correspondence analysis (CCA) biplots of representative water-quality variables in relation to fish community composition. [Refer to table 2 for definitions of variable abbreviations.]

variables and biotic communities, Whittier and others (1988) were able to readily discriminate montane from nonmontane ecoregions in Oregon. However, additional ordinations with biotic communities (fish, invertebrates, and periphyton) were necessary to discriminate among the nonmontane ecoregions. In the current study, ecoregions could not be distinguished, perhaps because there were only one or two sites within some of the ecoregions. Additionally, the lack of variability in the geomorphological characteristics from site to site may be responsible for not providing site discrimination or a sufficient gradient to be useful in ordination with the fish community data. The generally meandering form of these prairie streams and rivers does not change as the streams progress from headwaters to major rivers. Unlike streams in areas with substantially greater relief, the frequency of geomorphological units (riffles, pools, and runs) does not appear to change greatly as the streams increase in size. The most common geomorphological units are runs, which predominate in all sizes of streams. Fluvial processes, which act to modify the geomorphological composition of a stream (frequency and abundance of riffles, pools, and runs), have different effects in high-gradient upland streams compared to low-gradient meandering lowland streams. The differences in riffle and pool formation and frequency in high gradient streams compared to the propagation of meanders and runs in low gradient streams may account for the lack of significance of these geomorphological factors in site discrimination in the Red River Basin.

Part of the basic design of the NAWQA Program focused on land use. The Red River Basin is almost entirely agricultural. The terrestrial habitat characteristics were selected to reflect not only the abundance of agriculture in the Red River Basin, but also factors that relate to differences in cropping patterns (soil characteristics) and the differences among ecoregions (the relative amounts of wetlands and forested land). The spatial-data coverage based on aerial photographs and NWI maps was not sufficient to differentiate between cropland and pasture, a major distinction between some of the ecoregions, particularly the Northern Glaciated Plains and the Red River Valley. Differences between ecoregions were only apparent based on the amounts of forested land and riverine wetlands among sites. Even so, the sites did not group by ecoregions. Land use was examined at 300, 1,000, and 4,000 m distances from the upstream boundary and laterally from the reach boundary and the largest distance (4,000 m) was selected for analysis (Supplemental Information section). Quantification of land uses and land cover within this distance may not be sufficient to discern ecoregion differences, particularly when the dominant landscape feature of the area examined is a stream or river. The area examined was sufficient to discriminate land use among sites. In a similar GIS-based study of northern Minnesota tributaries to Lake Superior, Richards and Host (1994) were able to discern watershed land-use differences using a buffer area of only 100 m. They found that the land use characteristics of a 100 m buffer on either side of a stream were within five percent of the values determined for the entire watershed upstream of their sampling sites. The only exception they found was for wetlands which were more prevalent in the 100 m buffer than throughout the watershed.

The variable selection process identified variables relative to all three of the general terrestrial habitat characteristics of soils, land use/land cover, and riparian zone. Those variables selected through the process of cluster analysis and principle components analysis were judged to either have the most biological significance, or were most useful to discriminate among the sites. Selection of the 4,000 m distance was also judged to provide the best discrimination among sites within the limits of available data. Land use and riparian zone characteristics have been used to both rate and classify streams (Plafkin and others, 1989; Rankin, 1989; Peterson, 1992), and as such are considered to have a significant correlation with aquatic resource quality. The riparian zone has been shown to act as a buffer from various land disturbances including agriculture, grazing, mining, and urban construction (Waters, 1995) and so would be expected to have an effect on aquatic resource quality.

Several prior studies have found correlations of fish species assemblages with water-quality variables (Stevenson and others, 1974; Hawkes and others, 1986; Rohm and others, 1987; Whittier and others, 1988; Matthews and others, 1992). Although some of these studies focused on selected water-quality variables, the general conclusions indicated that except in extreme cases—for example pH related to acid mine drainage (Townsend and others, 1983)—water-quality parameters are interrelated as a multivariate system (Hawkes and others, 1986). Cluster analysis and PCA in this study also indicate that water-quality parameters tend to be interrelated. The process of identifying representative variables indicates that parameters related to ionic composition and concentration are correlated as are the nutrient parameters (Supplemental Information section).

The water-quality variables that were correlated with the first ordination axis in the CCA were the minimum percentage of fine suspended sediment and minimum concentration of dissolved oxygen. These variables also represented minimum concentrations of dissolved organic carbon, ammonia plus organic nitrogen, and dissolved calcium. The gradient expressed by this ordination axis shows clear-water streams compared to other more turbid streams which exhibit a larger range of ionic concentrations. In the Red River Basin, this gradient generally describes the difference between the more turbid streams in the Red River Valley ecoregion compared to less turbid streams outside that ecoregion (Fandrei and others, 1988). The second ordination axis is related to maximum nutrient (nitrogen and phosphorus) concentrations and may reflect the underlying agricultural development of the basin. It is important to note that the influence of agricultural chemicals as represented by the second axis of the ordination is less important in explaining the variability in fish community composition than the first axis. The majority of nutrients from fertilizers are carried to the streams during snowmelt runoff during early spring (Lan Tornes, USGS, written communication, 1996). At this time, water temperatures are just above freezing and little biological utilization of the nutrients occurs. The nutrients are flushed through the system into the mainstem of the Red River and eventually enter Lake Winnipeg north of the study unit where nutrient enrichment has been documented (Brunskill and others, 1980).

Hydrologic variability has been related to fish species composition (Poff and Ward, 1989; Poff and Allen, 1995). In streams of Minnesota and Wisconsin, Poff and Allen (1995) classified fish as associated with stability of hydrologic conditions. They defined streams as either stable or variable based the coefficient of variation of flow and predictability of flow. Many of the species they classified are found in the Red River Basin. Fish species including shorthead redhorse, smallmouth bass, longnose dace, and rosyface shiner were associated with stable sites. Species associated with variable streams included black bullhead and yellow perch. Perhaps more significant were the species that could not be distinctly associated with either a stable or variable hydrologic regime. These were white sucker, common shiner, creek chub, Johnny darter, bluntnose minnow, spotfin shiner, common carp, hornyhead chub, sand shiner, fathead minnow, northern pike, stone cat, and blackside darter. This group of fish corresponds to the most common species found during this study (table 4), with a few exceptions. This may indicate that hydrologic variability is not a major factor affecting the most common species in the Red River Basin, but may affect the other less common species.

The first axis of the CCA ordination correlated with high discharge variability in spring and also appeared to relate to river size. High variability in discharge would normally be associated with smaller streams. It may be that the coefficients of variation of mean monthly discharge act to smooth the variation in small stream discharge because they are more flashy. Alternative variables such as the coefficient of variation of daily discharge or the percent change in daily or monthly maximum and minimum flows might be a better measure of hydrologic variability.

Fish species richness tends to increase with stream size by the addition of species rather than replacement (Shelford, 1911; Sheldon, 1968). The ordination of fish community composition for streams in the Red River Basin agrees with this trend. Although the increase in species richness has been attributed to the addition of species (Shelford, 1911; Sheldon, 1968; Goldstein, 1981) it is also related to the number of ecoregions a river flows through (Goldstein, 1995). In the Red River Basin there appears to be a large-river ichthyofauna with some species only collected from the mainstem of the Red River or the larger tributaries, those sites initially designated as large nonwadeable sites. These species include silver lamprey, goldeye, mooneye, silver chub, river shiner, quillback, and sauger. Large-river species have been identified for large rivers in Missouri (Pfliger, 1971), Kentucky (Burr and Warren, 1986), and Indiana (Simon, 1992). There also appears to be a small-river ichthyofauna composed of longnose and blacknose dace, Johnny darter, Iowa darter, common shiner, bigmouth shiner, brook stickleback, tadpole madtom, creek chub, northern redbelly dace, and central mudminnow (fig. 9). Designation of large-river and headwater species for the Red River Basin has been proposed (Goldstein and others, 1994).

During the study, the climatic influences did not vary greatly across the Red River Basin. High amounts of precipitation occurred in all ecoregions. The zoogeography of fish of the Red River Basin is related to the immigration of fish species primarily from the Mississippi River (Underhill, 1989). The common source of species and the relatively short time since the last glaciation precludes distinction of sites by ecoregions based on a characteristic ichthyofauna as was possible in Oregon (Whittier and others, 1988). The similarity of species composition and abundance across the sites is reflected by the relatively small amount of variance explained by the species data in the DCA (approximately 34 percent on the first two axes). The common core species of the small streams tend to obscure differences and make interpretation of the ordinations more difficult.

It is apparent that in the Red River Basin, there is no single environmental factor that controls the composition and abundance of fish. Each of the environmental data sets contained representative variables that correlated to the axes in a similar manner. The original data analysis plan was to use CCA to correlate all the representative environmental variables with the fish community data. That was not possible due the lack of collocated sites and redundancy. Even so, it is doubtful that much greater resolution in fish community variation would have been achieved with a combined analysis. The understanding of the interrelationship among physical and chemical environmental factors has not yet reached the stage where prediction of cause and effect can be made with certainty and the resultant expression of that cause and effect in a biological community known.

At best, the water-quality data could be used to explain about one-third of the variance in the fish community composition (table 9). The other environmental variables were able to explain about 60

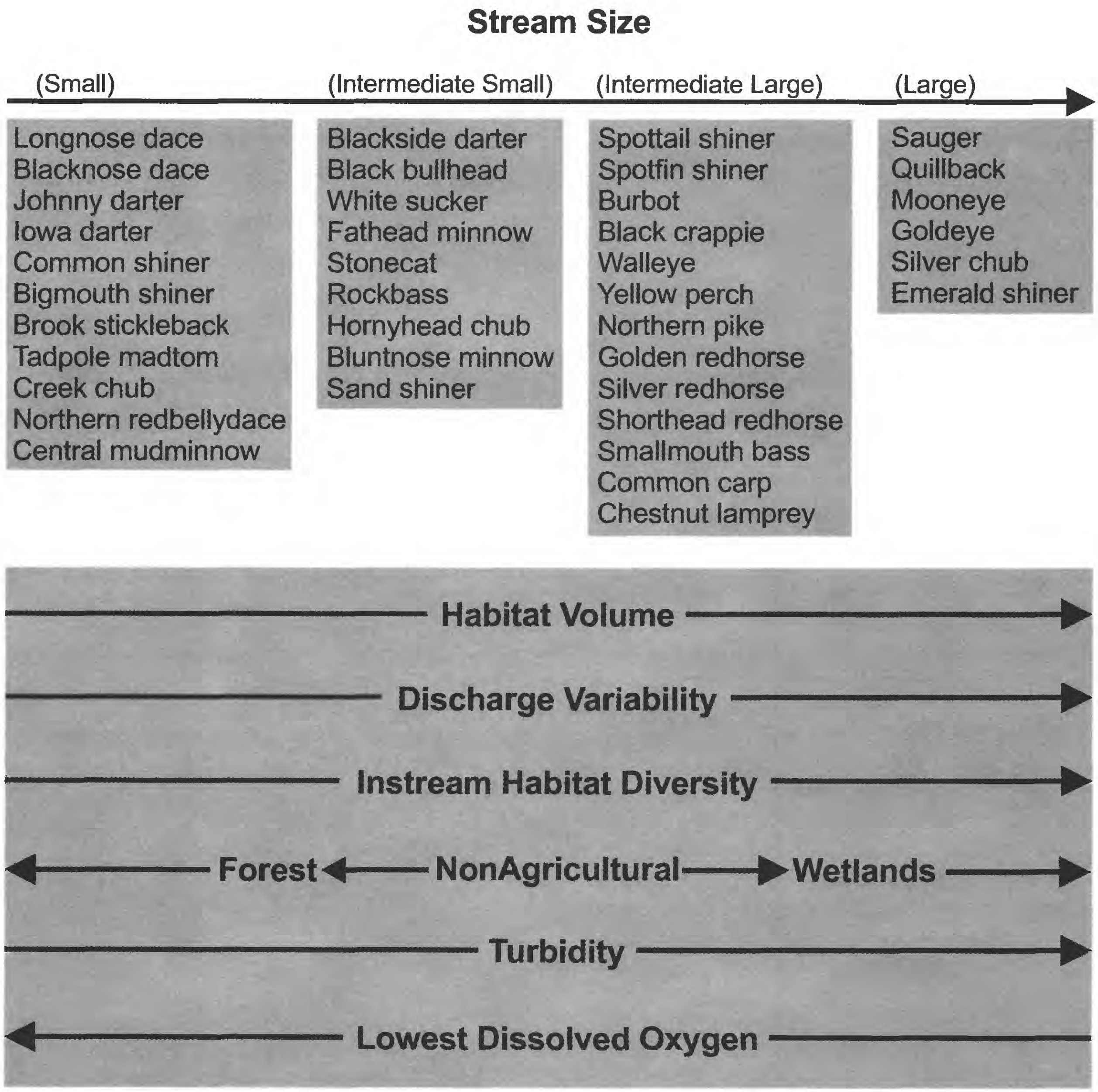

Figure 9. Distribution of fish species in the Red River of the North Basin in relation to environmental gradients.

percent of the variance in the fish community composition. The combined analysis of the physical and chemical variables could be used to explain much of the variance in the composition of the fish communities in the Red River Basin. The remaining variation comes from several sources. Most significant may be the biological interaction of species through predation and competition. However, without extensive studies of community structure, it would be very difficult to determine the consequences of biological interactions. Other sources are natural variation in abundance, environmental factors not measured, and anthropogenic sources of disturbance (chronic-toxic nonpoint, point, and atmospheric inputs).

Alternatively, the three years of unusually high streamflow may have had a significant effect on the fish communities. High streamflow may have altered the communities such that only flood-resistant species comprised the majority of the fish community. Many of the common species found during this study (common carp, creek chub, hornyhead chub, fathead minnow, common shiner, white sucker, northern pike, blackside darter, and johnny darter) were classified by Poff and

Allen (1995) as distinctly associated with neither variable nor stable streams. These species may be better adapted to resist high flows, and, therefore, could be the core species of this basin. Core species would be those species whose niche is sufficiently broad to be able to capitalize on disturbed environments. The fish community may not have had sufficient recovery time for the more rare and less resistant species to re-invade from refugia. If this were the case, the fish communities at all sites would be similar with some differences based on the pool of species in the process of recolonization.

Summary

The U.S. Geological Survey is conducting a National Water-Quality Assessment. The Red River of the North (Red River) Basin is one of 60 study units that comprise the assessment. One of the goals of the program is to provide a nationally consistent description of water-quality conditions by identifying, describing, and explaining major factors that affect observed conditions. The purpose of this report is to define physical and chemical factors influencing fish community composition in the Red River Basin.

Approximately 8,500 years ago, the Red River Basin held glacial Lake Agassiz. Fish migrated into the basin primarily from the Mississippi River drainage. After the lake drained, the lake sediments were highly productive soils, making the Red River Basin an important agricultural area. A land-classification system based on ecologically similar areas (depending on soils, climate, natural vegetation, and geology) defines six ecoregions in the United States part of the basin.

Ecoregions, land use, and stream size were used to select 22 sites for ecological sampling. At six sites, one in each of five ecoregions and one at the Canadian border, three reaches were established. A single reach was defined at the other sites. At each reach measurements were made of the physical habitat of the stream. A GIS-based system was used to determine segment and basin characteristics at sites where data were available on soils, land use/land cover, and the riparian zone. Data were derived from aerial photographs, National Wetlands Inventory data, soils maps from the Natural Resource Conservation Service, and the U.S. Department of Agriculture soils data base. Water-quality data were collected from some of the ecological sites and other sites within the basin. Hydrologic data were assembled from U.S. Geological Survey discharge records from throughout the basin. Not every type of data were available for every site.

Data analysis of physical and chemical data sets involved correlation analysis, cluster analysis, and principal components analysis. These procedures reduced the number of variables to a few representative variables. The representative variables were then analyzed with the fish community composition data by canonical correspondence analysis to determine which environmental gradients explain the variability in fish community composition.

Within the set of instream habitat measurements, stream size (channel width and depth) and habitat diversity (number of substrates, cover, variability of current) were the two gradients best correlated with fish community composition. Among the terrestrial habitat measures, nonagricultural land cover and riparian zone integrity were the gradients that best explained variation in fish community composition. Hydrologic variables related to stream discharge variability and the average annual frequencies of high and low discharges correlated to fish community composition. Fish community composition was correlated to gradients of turbidity, minimum concentrations of dissolved oxygen and organic carbon, and the range of concentrations of dissolved ions and nutrients.

In the Red River Basin, there are some distinct fish species associated with large rivers and others with small streams. The distribution of fish is influenced by their physical and chemical environment. No single environmental factor as measured during the study could be identified as the major factor influencing fish community composition. Rather, it appears that physical and chemical environmental factors act in combination to affect fish community composition.

References

Brady, N.C., 1984, The nature and properties of soils, (9th ed.): MacMillan Publishing Company, New York, 750 p.

Brunskill, G.J., Elliott, S.E.M., and Campbell, P., 1980, Morphometry, hydrology, and watershed data pertinent to the limnology of Lake Winnipeg: Canadian Manuscript Report of Fisheries and Aquatic Sciences 1556, 32 p.

Burr, B.M., and Warren, M.L., Jr., 1986, A distributional atlas of Kentucky fishes: Kentucky Nature Preserves Commission, Scientific and Technical Series Number 4, 398 p.

Cohen, Phillip, Alley, W.M., and Wilber, W.G., 1988, National water-quality assessment—Future directions of the U.S. Geological Survey: Water Resources Bulletin, v. 46, no. 6, p. 1147-1151.

Colwell, R.K., 1974, Predictability, constancy, and contingency of periodic phenomena: Ecology, v. 55, p. 1148-1153.

Fandrei, Gary, Heiskary, Steve, and McCollar, Sylvia, 1988, Descriptive characteristics of the seven ecoregions in Minnesota: Minnesota Pollution Control Agency, 140 p.

Frissell, C.A., Liss, W.J., Warren, C.E., and Hurley, M.D., 1986, A hierarchical framework for stream habitat classification—Viewing streams in a watershed context: Environmental Management, v. 10, p. 199-214.

Gauch, H.G., 1982, Multivariate analysis in community ecology: Cambridge University Press, Cambridge, Mass., 298 p.

Goldstein, R.M., 1981, Longitudinal succession in impact assessment of river system fish communities: Water Resources Bulletin, v. 17, p. 75-81.

Goldstein, R.M., 1995, Aquatic communities and contaminants in fish from the streams of the Red River of the North Basin, Minnesota and North Dakota: U.S. Geological Survey Water-Resources Investigations Report 95-4047, 34 p.

Goldstein, R.M., Simon, T.P., Bailey, P.A., Ell, Michael, Pearson, Eric, Schmidt, Konrad, and Enblom, J.W., 1994, Concepts for an index of biotic integrity for streams of the Red River of the North Basin: Proceedings of the North Dakota Water Quality Symposium, March 30-31, 1994, Fargo, North Dakota, p. 169-180

Gorman, O.T., and Karr, J.R., 1978, Habitat structure and stream fish communities: Ecology, v. 59, p. 507-515.

Green, R.H., and Vascotto, G.L., 1978, A method for the analysis of environmental factors controlling patterns of species composition in aquatic communities: Water Research, v.12, p. 583-590.

Gurtz. M.E., 1994, Design of the biological components of the National Water-Quality Assessment (NAWQA) Program, *in* Loeb, S.L. and Spacie, Anne, eds., Biological monitoring of aquatic systems: CRC Press, Lewis Publishers, Boca Raton, Florida, p. 323-354.

Harkness, R.E., Haffield, N.D., Berkas, W.R., and Norbeck, S.W., 1994, Water-resources data for North Dakota, water year 1993: U.S. Geological Survey Water-Data Report ND-93-1, 450 p.

Hawkes, C.L., Miller, D.L., and Layher, W.G., 1986, Fish ecoregions of Kansas—Stream fish assemblage patterns and associated environmental correlates: Environmental Biology of Fishes, v. 17, p. 267-279.

Helsel, D.R., and Hirsh, R.M., 1992, Statistical methods in water resources: Elsevier, New York, 522 p.

Horwitz, R.J., 1978, Temporal variability patterns and the distributional patterns of stream fishes: Ecological Monographs, v. 48, p. 307-321.

International Red River Pollution Board, 1990, Thirtieth Progress Report to the International Joint Commission—Red River: International Red River Pollution Board, October 1990, 115 p.

Leahy, P.P., Rosenshein, J.S., and Knopman, D.S., 1990, Implementation plan for the National Water-Quality Assessment Program: U.S. Geological Survey Open-File Report 90-174, 10 p.

Lorenz, D.L., and Stoner, J.D., 1996, Sampling design for a water-quality assessment of the Red River of the North Basin, Minnesota, North Dakota, and South Dakota, 1993-95: U.S. Geological Survey Water-Resources Investigation Report 96-4129, 2 plates, 1:1,500,000 and 1:750,000.

Lotspeich, F.B., and Platts, W.S., 1982, An integrated land-aquatic classification system: North American Journal of Fisheries Management, v. 2, p. 138-149.

Maclay, R.W., Winter, T.C., and Bidwell, L.E., 1972, Water resources of the Red River of the North Basin, Minnesota: U.S. Geological Survey Water-Resources Investigations 72-1, 127 p.

Matthews, W.J., Hough, D.J., and Robison, H.W., 1992, Similarities in fish distribution and water quality patterns in streams of Arkansas—Congruence of multivariate analyses: Copeia, v. 1992, p. 296-305.

Meador, M.R., Hupp, C.R., Cuffney, T.F., and Gurtz, M.E., 1993a, Methods for characterizing stream habitat as part of the National Water-Quality Assessment Program: U.S. Geological Survey Open-File Report 93-408, 48 p.

Meador, M.R., Cuffney, T.F., and Gurtz, M.E., 1993b, Methods for sampling fish communities as part of the National Water-Quality Assessment Program: U. S. Geological Survey Open-File Report 93-104, 40 p.

Menge, B.A., and Sutherland, J.P., 1976, Species diversity gradients—Synthesis of the roles of predation, competition, and temporal heterogeneity: American Naturalist, v. 110, p. 351-369.

Minnesota Soil Survey Staff, 1983, Soil associations of Minnesota: Minnesota Agricultural Experiment Station, University of Minnesota, St. Paul, 1 plate.

Mitton, G.B., Hess, J.H., and Guttormson, K.G., 1994, Water-resources data for Minnesota, water year 1993, v. 1., Great Lakes and the Souris-Red-Rainy River Basins: U.S. Geological Survey Water-Data Report MN-93-1, 230 p.

Murphy, M.L., Hawkins, C.P., and Anderson, N.H., 1982, Effects of canopy modification and accumulated sediment on stream communities: Transactions of the American Fisheries Society, v. 110, p. 469-478.

Omernik, J.A., and Gallant, A.L., 1988, Ecoregions of the upper midwest states: U.S. Environmental Protection Agency, EPA/600/3-88/037, 56 p.

Omodt, H.W., Johnsgard, G.A., Patterson, D.D., and Olson, O.P., 1968, The major soils of North Dakota: North Dakota State University Agricultural Experiment Station Bulletin 472, 60 p., 1 map.

Petersen, R.C., Jr., 1992, The RCE—A riparian, channel, and environmental inventory for small streams in the agricultural landscape: Freshwater Biology, v. 26, p. 295-306.

Pflieger, W.L., 1971, A distributional study of Missouri fishes: University of Kansas Museum of Natural History Number 20, p. 225-570.

Plafkin, J.L., Barbour, M.T., Porter, K.D., Gross, S.K., and Hughes, R.M., 1989, Rapid bioassessment protocols for use in streams and rivers—benthic macroinvertebrates and fish: U.S. Environmental Protection Agency Assessment and Watershed Protection Division, EPA/440/4-89/001, variously paged.

Poff, N.L., and Allen, J.D., 1995, Functional organization of stream fish assemblages in relation to hydrological variability: Ecology, v. 76, p. 606-627.

Poff, N.L., and Ward, J.V., 1989, Implications of streamflow variability and predictability for lotic community structure—A regional analysis of streamflow patterns: Canadian Journal of Fisheries and Aquatic Sciences, v. 46, p. 1805-1818.

Power, M.E., Stout, R.J., Cussing, C.E., Harper, P.P., Hauer, F.R., Matthews, W.J., Moyle, P.B., Statzner, Bernhard, and Wais De Badgen, I.R., 1988, Biotic and abiotic controls in river and stream communities: Journal of the North American Benthological Society, v. 7, p. 456-479.

Rankin, E.T., 1989, The qualitative habitat evaluation index (QHEI)—Rationale, methods, and application: State of Ohio Environmental Protection Agency, Ecological Assessment Section Division of Water Quality and Assessment, variously paged.

Reice, S.R., Wissmar, R.C., and Naiman, R.J., 1990, Disturbance regimes, resilience, and recovery of animal communities and habitats in lotic ecosystems: Environmental Management, v.14, p. 647-660.

Richards, Carl, and Host, G.E., 1994, Examining land use influences on stream habitats and macroinvertebrates—A GIS approach: Water Resources Bulletin, v. 30, p. 729-738.

Ricklefs, R.E., 1987, Community diversity—relative roles of local and regional processes: Science, v. 235, p. 167-171.

Rohm, C.M., Giese, J.W., and Bennett, C.C., 1987, Evaluation of an aquatic ecoregion classification of streams in Arkansas: Journal of Freshwater Ecology, v. 4, p. 127-140.

SAS Institute Inc., 1989, SAS/STAT User's guide, Version 6, fourth edition, vol 1: SAS Institute Inc., Cary, North Carolina, 943 p.

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

Schlosser, I.J., 1987, The role of predation in age and size related habitat use by fishes: Ecology, v. 68, p. 651-659.

Sheldon, A.L., 1968, Species diversity and longitudinal succession in stream fishes: Ecology, v. 49, p. 193-198.

Shelford, V.A., 1911, Ecological succession. I. Stream fishes and the method of physiographic analysis: Biological Bulletin, v. 21, p. 9-35.

Shelton, L.R., 1994, Field guide for collecting and processing stream-water samples for the National Water-Quality Program: U.S. Geological Survey Open-File Report 94-455, 42 p.

Simon, T.P., 1992, Biological criteria development for large rivers with an emphasis on an assessment of the White River drainage, Indiana: U.S. Environmental Protection Agency, Ambient Monitoring Section, Chicago, Ill., EPA 905/R-92/006, variously paged.

Stevenson, M.M., Schnell, G.D., and Black, R., 1974, Factor analysis of fish distribution patterns in western and central Oklahoma: Systematic Zoology, v. 23, p. 202-218.

Stoner, J.D., 1991, National Water-Quality Assessment Program—Red River of the North: U.S. Geological Survey Open-File Report 91-151, 2 p.

Stoner, J.D., Lorenz, D.L., Wiche, G.J., and Goldstein, R. M., 1993, Red River of the North Basin, Minnesota, North Dakota, and South Dakota: Water Resources Bulletin, v. 29, p. 575-615.

Strahler, H.N., 1957, Quantitative analysis of watershed geomorphology: American Geophysical Union Transactions, v. 33, p. 913-920.

Ter Braak. C.J.F., 1986, Canonical correspondence analysis—A new eigenvector technique for multivariate direct gradient analysis: Ecology, v. 67, p. 1167-1179.

Ter Braak, C.J.F., 1987, CANOCO—A Fortran program for canonical community ordination by (partial) (detrended) (canonical) correspondence analysis, principal components analysis, and redundancy analysis, version 2.0 edition: TNO Institute of Applied Computer Science, Statistical Department Wageningen, 6700 AC Wageningen, The Netherlands, 95 p.

Tornes, L.H., and Brigham, M.E., 1994, Nutrients, suspended sediment, and pesticides in waters of the Red River of the North Basin, Minnesota, North Dakota, and South Dakota, 1970-90: U.S. Geological Survey Water-Resources Investigations Report 93-4231, 62 p.

Townsend, C.R., Hildrew, A.G., and Francis, Jeanette, 1983, Community structure in some southern English streams—The influence of physicochemical factors: Freshwater Biology, v. 13, p. 521-544.

Underhill, J.C., 1989, The distribution of Minnesota fishes and late Pleistocene glaciation: Journal of the Minnesota Academy of Science, v. 55, no.1, p. 32-37.

U.S. Department of Agriculture, 1975, Soil survey of Marshall County, South Dakota: United States Department of Agriculture, Soil Conservation Service and United States Department of Interior, Bureau of Indian Affairs, South Dakota Agricultural Experiment Station, 116 p.

U.S. Department of Agriculture, 1977a, Soil survey of Roberts County, South Dakota: United States Department of Agriculture, Soil Conservation Service and United States Department of Interior, Bureau of Indian Affairs, South Dakota Agricultural Experiment Station, 116 p.

U.S. Department of Agriculture, 1977b, Soil survey of Trail County, North Dakota: United States Department of Agriculture, Soil Conservation Service and North Dakota Agricultural Experiment Station, 140 p.

U.S. Department of Agriculture, 1991, State soil geographic data base (STASGO) data users guide:

Soil Conservation Service Miscellaneous Publication Number 1492, 88 p.

Waters, T.F., 1977, The streams and rivers of Minnesota: University of Minnesota Press, Minneapolis, Minnesota, 373 p.

Waters, T.F., 1995, Sediment in streams, sources, biological effects and controls: American Fisheries Society Monograph 7, 251 p.

Werner, E.E., Hall, D.J., Laughlin, D.R., Wagner, D.J., Wilsmann, L.A., and Funck, F.C., 1977, Habitat partitioning in a freshwater fish community: Journal of the Fisheries Research Board of Canada, v. 34, p. 360-370.

Whittier, T.R., Hughes, R.M., and Larson, D.P., 1988, Correspondence between ecoregions and spatial patterns in stream ecosystems in Oregon: Canadian Journal of Fisheries and Aquatic Sciences, v. 45, p. 1264-1278.

Winter, T.C., Benson, R.D., Engberg, R.A., Wiche, G.J., Emerson, D.G., Crosby, O.A., and Miller, J.E., 1984, Synopsis of ground-water and surface-water resources of North Dakota: U.S. Geological Survey Open-File Report 84-732, 127 p.

Wright, H.E., Jr., 1972, Quaternary history of Minnesota, *in* Sims, P.K., and Morey, G.B., ed., Geology of Minnesota—A centennial volume: Minnesota Geological Society, St. Paul, Minnesota, p. 515-560.

Supplemental Information

This section provides the details of the selection of representative variables from four data sets: instream habitat, terrestrial habitat, hydrology, and water quality. The objective of these analyses was to reduce the number of environmental variables to a number suitable for use in a CCA with the fish community composition data. Although the analytical approach was similar for each data set, there were differences in the specifics of the analyses. These differences were due to such factors as autocorrelation of variables, time periods when data were collected, multiple analysis of different forms of the data, and various transformations.

Included in this section is an analysis of the fish community composition data using cluster analysis and DCA. DCA scores were used to confirm the selection of representative environmental variables.

Instream Habitat

The first step was to summarize the instream habitat data (table 2) by use of descriptive statistics. The means, standard deviations, and coefficients of variation were calculated for the following variables: channel width, depth, channel width to depth ratio, stream flow velocity, canopy angle, reach gradient, reach sinuosity, and percent area with instream cover. Descriptions or ratings were used for bank erosion, bank shape, and bank stability (Meador and others, 1993a). The variables produced were frequency of erosion (number of observations of erosion at each transect/total number of transects), frequency of bank shapes which were not linear (bank shape at each transect was categorized as either concave, convex, or linear), the sum of the bank stability ratings (scored 1 to 4 depending on the amount of stabilizing vegetation with 4 the highest rating).

Several transformations were used to make certain data sets more normally distributed prior to statistical analysis. An arcsine transformation was applied to the following variables: erosion frequency, sum of stability indices, and the frequency of non-linear bank shapes. A logarithmic transformation was used for reach gradient. After the transformations were made, a correlation matrix was generated to determine which variables were correlated. This was done to remove redundancy. If two variables were significantly correlated ($r>0.6$, $p<0.05$), then they were in effect measuring the same habitat feature and one was removed from further analysis. A second correlation matrix was produced using the raw data (no transformations). Correlation coefficients from the two matrices were compared. No difference was observed between the correlation coefficients of transformed and untransformed data. Therefore in the following multivariate analyses, untransformed data were used. In the multivariate analyses, the data were standardized such that each variable had a mean of 0 and a standard deviation of 1. The standardization procedure (SAS, 1989) was used to equalize the importance of each variable, regardless of its magnitude or variance.

The correlation matrix of the instream habitat data indicated very few significant correlations among the variables. There were positive correlations among mean canopy angle, mean channel width, and mean channel depth. The mean width/depth ratio correlated with both mean depth and mean width and the other variables that contain a measure of stream size, and was deleted from other analyses. Velocity and sinuosity both decreased with the size measures. Instream cover increased with sinuosity. Stream bank stability and frequency of erosion were negatively correlated.

The instream habitat measures were sufficient to group streams by similar physical features. The groups determined by cluster analysis were not the same as the size groups initially established by stream basin area (table 1), nor did they conform to ecoregion boundaries.

Five main groups of streams were identified by the cluster analysis (fig. 10). Initial inspection indicated that size (mean channel width, mean depth, mean canopy angle, etc.) may be the dominant factor used for separation because Group 5 consisted primarily of the larger nonwadeable sites (table 1) and Group 4 contained the two smallest stream sites (as measured by channel width), the Turtle and Snake Rivers. Groups 1, 2, and 3 contained all the other rivers. The multiple reach sites, which were located in different ecoregions, separated out in the cluster analysis, one multiple reach site per cluster group. The only exception was the first reach on the Sheyenne River which clustered in Group 2 while the other two reaches clustered together in Group 3.

Although the multiple reach sites were all grouped by ecoregion or size, the other rivers did not consistently follow the ecoregion or size grouping. Groups 1, 2, and 3 each contain one multiple reach site. The other rivers that grouped with the multiple reach sites were not necessarily from the same ecoregion or the same size class. For example, in Group 2, the Goose and Sheyenne Rivers are not in the same ecoregion (Red River Valley and Northern Glaciated Plains, respectively) as the Roseau River (Northern Minnesota Wetlands). Both the Goose and the Sheyenne Rivers are mid-size wadeable streams while the Roseau River is a large nonwadeable river. The two multiple reach sites in the Red River

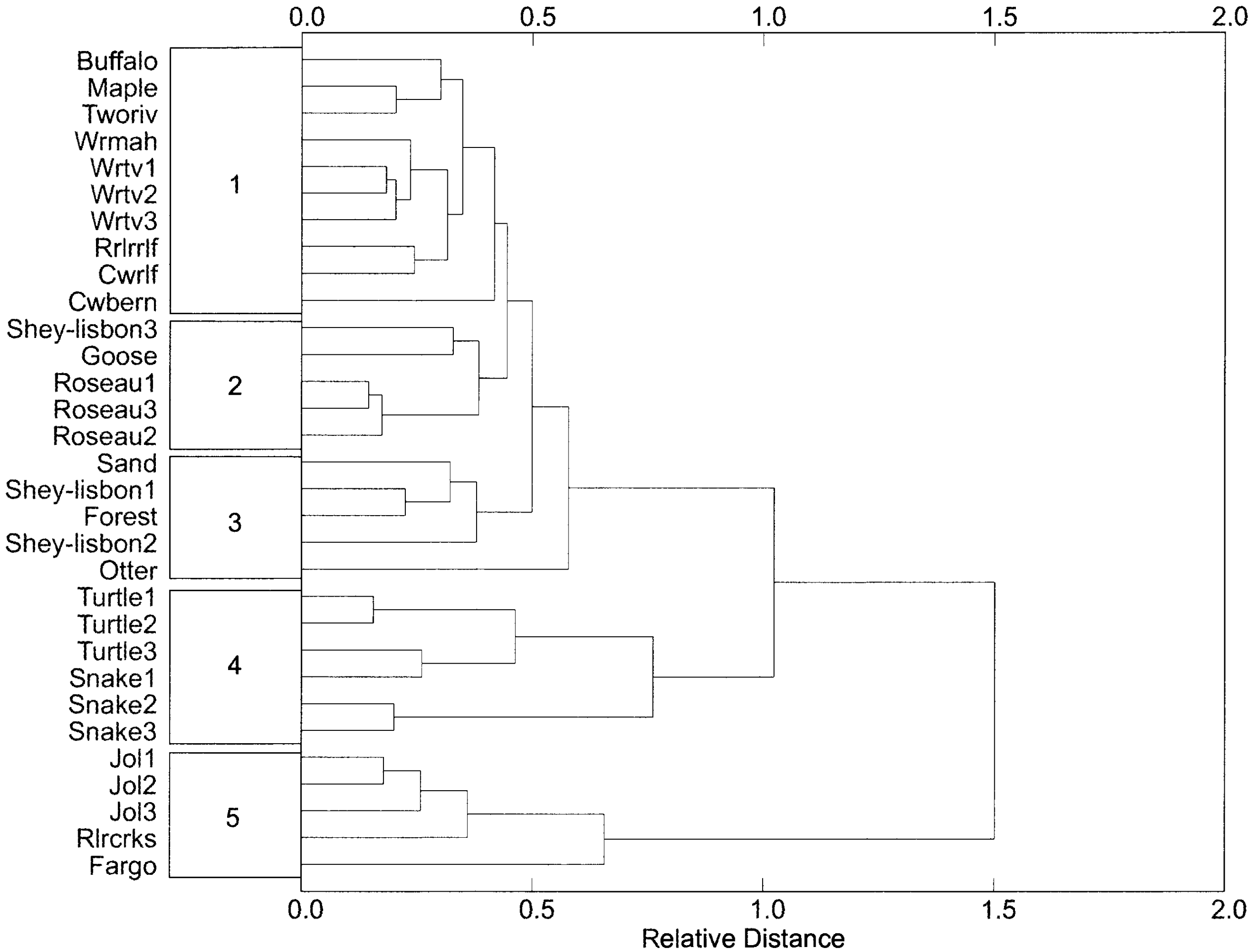

Figure 10. Cluster analysis dendrograph of ecological sites grouped by instream habitat measurements. [Refer to table 1 for definitions of site abbreviations.]

Valley ecoregion, the Snake River on the east side and the Turtle River on the west side, however, did group together.

PCA was used to determine which physical factors were most important for distinguishing the sites and explaining variability between and among sites. The eigenvectors (loadings) for the habitat variables (table 10) were relatively low with maximum absolute values approximately 0.40. The cumulative variance explained by the first three principal component axes was less than 60 percent. The major factors identified by the analysis for principal component Axis 1 are related primarily to size and geomorphological variability: mean width, mean depth, and the coefficient of variation of stream flow velocity. The variability of instream cover and stability of the stream banks appear to be the second and third principal components because the loadings were highest for variables which indicate variability of the canopy angle, the number of different substrates in the stream, the amount of instream cover, bank stability, and erosion frequency.

The loadings of the important variables determined from the PCA were used as input in an average linkage cluster analysis to verify the groupings of the variables. Four groups were identified in the cluster analysis (fig. 11). The first group contained the stream size related variables: mean channel width, mean depth, mean velocity, and mean canopy angle. The second group contained variables related to the availability of instream fish habitat: bank stability, which is indicative of overhanging vegetation, number of substrates, and amount of cover. The third group contained measurements related to the variability of the stream habitat: the coefficients of variation for depth, velocity, and canopy angle as well as the frequency of a linear shape on the right bank. The last group contained the

Table 10.—Summary of principal components analysis (PCA) of instream habitat variables.
[Only variables with at least one loading (absolute value) greater than 0.20 are reported; CV, coefficient of variation]

Variable	Principal component 1 loadings	Principal component 2 loadings	Principal component 3 loadings
Mean depth	-0.4000	-0.1075	0.0515
Mean channel width	-.3617	.0047	.1325
CV velocity	.3475	.2583	-.1027
Mean canopy angle	-.2976	.2824	.0387
Mean velocity	-.2807	.2583	-.1027
CV depth	.1363	-.2640	-.1713
CV canopy angle	.2865	-.3582	-.1409
Bank stability left bank	.1937	.2721	.3205
Bank stability right bank	.2396	.0371	.3839
Frequency of erosion left bank	.1307	.2957	-.4176
Frequency of erosion right bank	.0887	.3637	-.2762
Frequency of linear bank shape-left bank	.2157	.2498	-.2764
Frequency of linear bank shape-right bank	.2836	.0382	.0196
Number of different substrate types	.1037	.3971	.3430
Percent area with instream cover	.1296	.1594	.3908
Reach gradient	.1373	.2901	-.2430
Summary			
Eigenvalue	4.6545	3.1179	2.0784
Proportion of total variance	.2738	.1834	.1223
Cumulative proportion	.2738	.4572	.5795

remaining variables related to bank instability: frequency of erosion, reach slope, and the frequency of left bank shape as linear.

One representative variable was selected from each cluster to be used as input to the CCA. The following variables were selected: mean channel width, number of substrates, coefficient of variation of flow velocity, and erosion frequency. Erosion frequency was the combined frequency from both the left and right banks and was used because individual bank erosion frequencies clustered in the same group. Each of these selected variables are considered to represent the other variables in its respective cluster group.

Terrestrial Habitat

Segment level data analysis followed the same approach as reach level data. Summary statistics were determined for the 300, 1,000, and 4,000 m distance classes. A correlation matrix was constructed to identify redundancy and the final data set standardized.

Within the correlation matrix, the high wind erodibility class was significantly correlated ($p<0.001$) with the medium water erosion class ($r=0.87$), and the high and low soil permeability classes ($r=0.91$ and 0.88, respectively). Therefore, this variable was removed from the data set. Land-use variables for the percent of water and wetlands were removed because they were redundant with the percentage of wetlands classified as lacustrine, riverine, and palustrine by the NWI coverage.

Average linkage cluster analysis was applied to each distance class to determine which distance class produced the best discrimination among the sites. Although all three cluster analyses were able to discriminate among the sites, the best groupings were from the 4,000 m distance class, the largest of the three. There were two main clusters and two sites that were outliers and did not cluster with the other sites (fig. 12). The data from this distance class were used in the

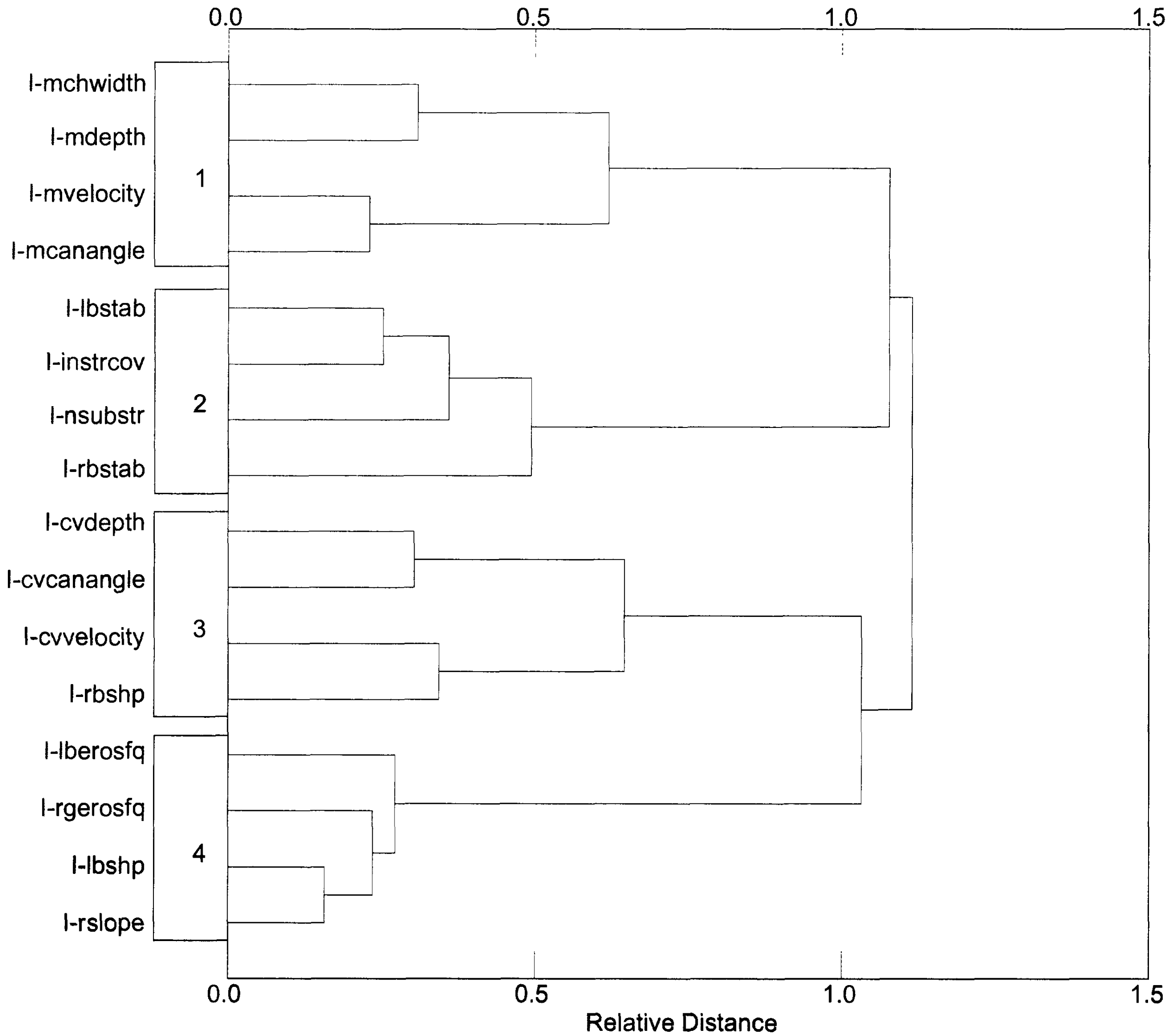

Figure 11. Cluster anaylsis dendrograph of instream habitat variable loadings from principal components analysis (PCA). [Refer to table 2 for definitions of variable abbreviations.]

analysis of landscape features and the correlations to the fish communities.

The first cluster group included only Minnesota sites. The second group included all the North Dakota sites and one site in Minnesota, the Wild Rice River at Mahnomen. The two sites that did not group with any other sites were the Clearwater River at Berner and the Goose River at Hillsboro, a Minnesota and a North Dakota site, respectively. At both these sites, at least one bank is almost devoid of trees. In the case of the Clearwater River at Berner, the stream has been channelized and riparian trees have not yet grown. At the Goose River site, one bank is part of a city park, and has very few trees. The sites did not cluster based on ecoregion or physiographic area even though soils and land use are major factors in the definition of both ecoregions and physiographic area.

The PCA indicated that soil factors were most important for grouping sites (table 11). On the first principal component axis, six of the eight highest loadings (>0.25) were from the soils classification groups. The highest loadings were the high water erosion class (0.3181) and the low permeability class (0.3193). The high classes had positive loadings while

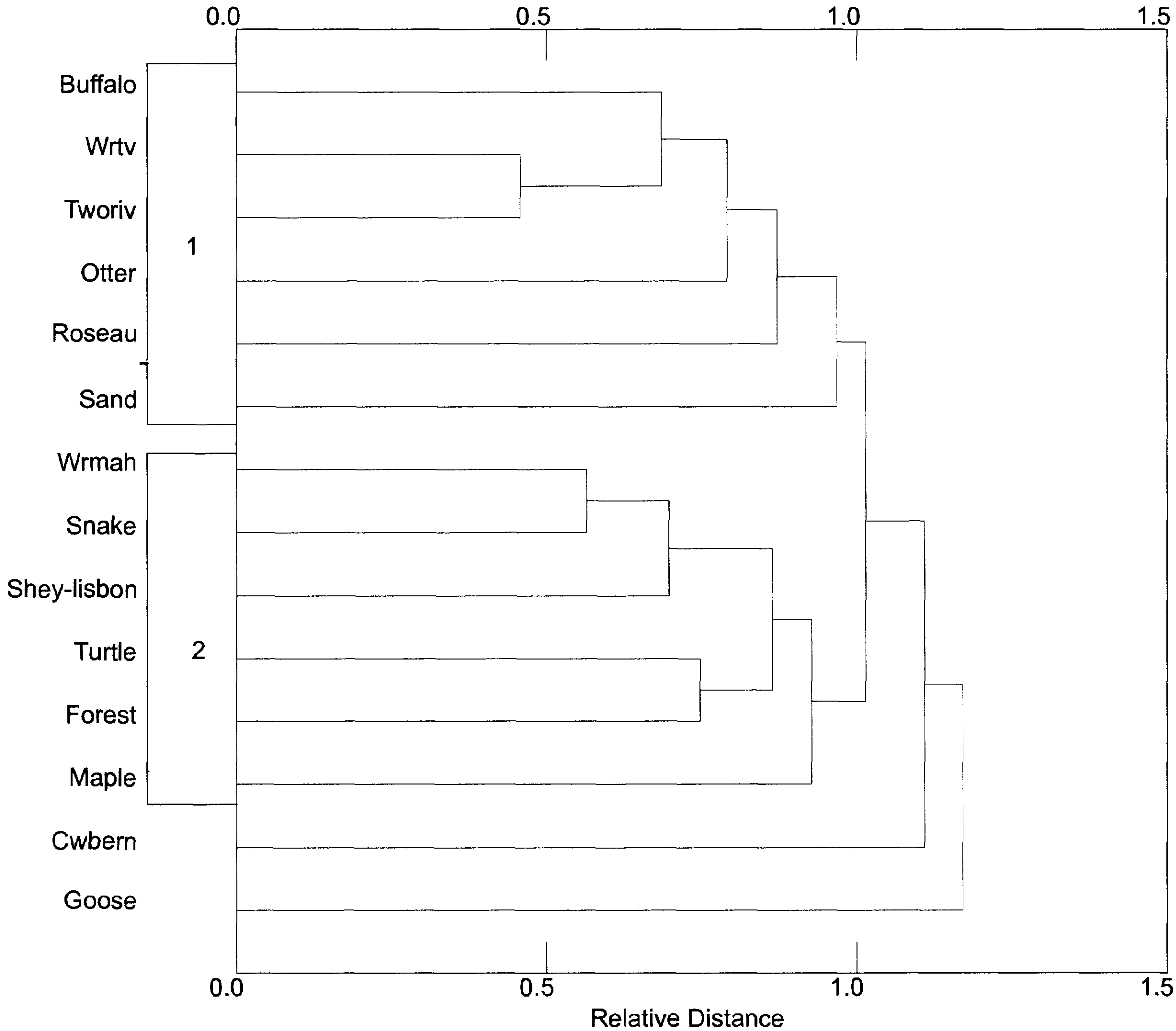

Figure 12. Cluster analysis dendrograph of ecological sites grouped by terrestrial habitat characteristics. [Refer to table 1 for definitions of site abbreviations.]

the low or moderate classes had strong negative loadings (table 11). The only other variables with high loadings on the first axis were left bank riparian zone discontinuity (0.2581) and the number of tributaries entering within 4,000 m of the upstream boundary of the reach (0.2520). The riparian zone variables were most important on the second principal component axis. Right bank riparian zone discontinuity, mean left bank riparian zone width, and coefficient of variation of right bank riparian zone width all exhibited high loadings (0.3879, 0.3013, and 0.3083; respectively). Other terrestrial habitat features that had high loadings on the second axis include the relative amount of forest land (0.2850), and the relative amount of palustrine wetland. These two variables help to explain the distinction of sites along state boundaries because generally the Minnesota sites were more forested than the North Dakota sites, and the Minnesota portion of the Red River Basin contains more wetlands than the North Dakota portion. One soils classification variable was still important on the second axis—the relative amount of soil in the moderate wind erosion class which had a loading of 0.3365.

On the third principal components axis, the highest loading was associated with riparian zone tree density

Table 11. Summary of principal components analysis (PCA) of terrestrial habitat characteristics.
[Only variables with at least one loading (absolute value) greater than 0.25 are reported; CV, coefficient of variation].

Basin and segment variable	Principal component 1 loadings	Principal component 2 loadings	Principal component 3 loadings
Shallow permeability-low	0.3193	0.1900	0.0515
Wind erodibility-high	.3181	.1606	.1204
Wind erodibility-moderate	-.2935	-.0156	-.0165
Shallow permeability-high	-.2780	-.1261	.0615
Shallow permeability-moderate	-.2684	.1307	.0468
Wind erodibility-low	.2593	-.1204	-.0339
Bank riparian discontinuity-left	.2581	-.153	-.2100
Number of tributaries	.2520	.2083	.0925
Bank riparian discontinuity-right	.0715	.3879	-.2636
Wind erodibility-moderate	.1884	.3365	.1071
Mean width of riparian zone-left bank	-.2083	.3103	.0475
CV riparian zone width-right bank	-.0350	.3083	-.1815
Palustrine wetlands	-.0864	.3054	.0653
Forested	-.2137	.2850	.2133
Tree density of riparian zone-right bank	.1570	-.0970	.4115
Riverine wetlands	.0136	-.0988	.3861
Lacustrine wetlands	.1667	-.1300	.3639
Number of downstream dams	.0727	-.2123	-.2832
Stream gradient	-.0546	-.1173	-.2808
CV left bank riparian zone width-left bank	.1945	-.0217	-.2533
Summary			
Eigenvalue	7.3258	3.8666	3.4909
Proportion of total variance	.2616	.1381	.1247
Cumulative proportion	.2616	.3997	.6257

on the right bank (table 11). The next two highest loadings were from the wetlands classification. The relative amounts of riverine and lacustrine wetlands had eigenvectors of 0.3861 and 0.3639, respectively. Relatively high loadings (> 0.25) were associated with the number of downstream dams within 6.2 km of the reach (-0.2832), stream gradient (-0.2808), right bank riparian zone discontinuity (-0.264), and the coefficient of variation of the left bank riparian zone width (0.2533).

A cluster analysis of the loadings (absolute value >0.25) from the principal components analysis indicated four groups of variables (fig. 13). None of the clusters contained variables exclusively that had high loadings from a single axis. The first cluster contained the number of downstream dams, stream gradient, left bank riparian discontinuity, the coefficient of variation of left bank riparian zone width, and the percentage of soils in the low wind erodibility class. The second cluster contained three soils classifications: high erodibility, low permeability, and moderate wind erodibility. The second cluster also contained the number of tributaries and right bank width coefficient of variation and riparian discontinuity. The third cluster also contained three soils classifications: moderate erodibility, and moderate and high permeability. Mean left bank riparian zone width, percent of forested land, and percent of palustrine wetlands complete the third cluster. The fourth cluster group contained the other two other wetlands types and the density of trees on the right bank.

One variable from each of the four clusters was

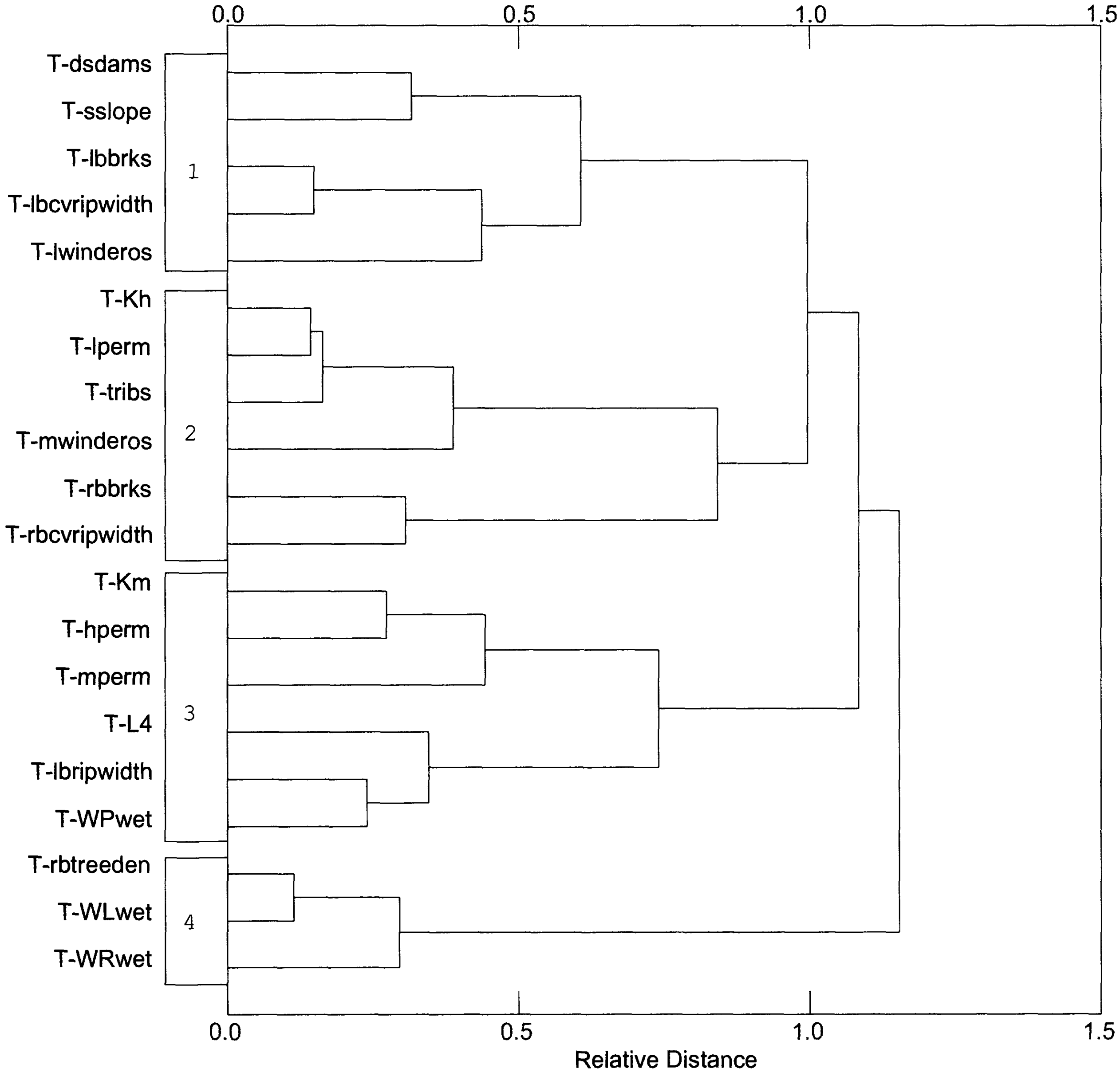

Figure 13. Cluster analysis dendrograph of terrestrial habitat variable loadings from principal components analysis (PCA). [Refer to table 2 for definitions of variable abbreviations.]

selected as the representative variable for use in the CCA. The selected variables were left bank discontinuity, percent of soils classified as highly erodible, percent of forested land, and percent of riverine wetlands.

Hydrology

Hydrologic variability data were treated in a similar manner to the other data sets. A correlation matrix of mean monthly flow for the period of record was constructed. All mean monthly flows correlated highly ($r > 0.82$, $p < 0.001$). These variables were excluded from further analyses.

Next, coefficients of variation of mean monthly flow were examined in a correlation matrix. Then a cluster analysis was applied to the coefficients of variation of mean monthly flow to group months with similar amounts and patterns of variation. A single month was selected to represent a cluster group (if more than one month was in the cluster). The month selected was the month that had the highest correlation coefficients with the other months in the cluster.

The correlation matrix of the coefficients of variation of mean monthly flow showed high correlations (r>0.8, p<0.001) between many of the months. The most significant correlations occurred for the fall and winter months (September through February). Significant correlations most often occurred between a month and the preceding month or the following month. The cluster analysis of the coefficients of variation of mean monthly discharge identified three clusters with more than one month in the group and three single months that did not group well with the other clusters. The first cluster group of months contained the spring months of March, April, May, and June. April was selected as the representative spring month because the three correlation coefficients for April with the other spring months were greater than any other spring month (r=0.70, March; r=0.62, May; and r=0.72, June; p<0.002).

The second monthly group was early winter: October, November, and December. October was selected as the representative month. Correlation coefficients with October were: r=0.91, November r=0.88, and December; p<0.001. Late winter months January and February grouped together. January was arbitrarily selected as the representative month. The other three months, July, August, and September, did not group well with other months. Coefficients of variation of mean monthly flow for these three months were dissimilar to any of the other months of the year.

Mean high and low flow values for 1, 3, 7, 30, and 90 days were examined in a correlation matrix. All high flow values correlated (r>0.99, p<0.001) as did all low flow values (r>0.98, p<0.001). One value of each was selected for inclusion in additional analyses. The seven-day high and low flows were selected because a seven day period of extreme conditions (high or low) should have biological relevance. A seven day period would convey prolonged low-flow implications while still being able to reflect high-flow conditions which occurred during the shorter time periods.

A correlation matrix that contained all the remaining variables and the reduced list from above was constructed to determine any additional redundancy. In addition to the coefficients of variation of mean monthly flow for April, October, January, July, August, September, and the mean seven-day high and low flows, the following variables were included: frequency and duration of low flows (25th percentile), frequency and duration of high flows (75th percentile), mean Julian date of annual high and low flow, mean annual duration of zero flow, and mean annual one-day high and low flow values. The mean annual one-day high and low flow values correlated highly (r>0.94, p<0.001) with the mean seven-day high and low flow values and so were removed from the list of variables used for site distinction.

The cluster analysis of all 33 sites produced nine clusters and several outliers that did not cluster with any group. The cluster groups contained two to six sites. The cluster groups tended to reflect geographical similarity. For example, one cluster contained the Sheyenne River sites at Kindred and West Fargo and the Red River sites at Whapeton-Breckenridge and Fargo. Another cluster group contained two sites on the Buffalo River, two sites on the Wild Rice River (Minnesota), and one site on the Roseau River. All these sites were in a band extending northeast from the central part of the Minnesota half of the basin. The other discriminating factor appears to be the presence of an upstream impoundment near the site. The sites that had an upstream impoundment (Otter Tail River at Fergus Falls, Maple River at Enderlin, and Red Lake River at Red Lake) were outliers and did not group with any of the other sites.

The PCA of the hydrologic variables accounted for 68 percent of the variance in the data (table 12). On the first axis, the coefficients of variation of the mean monthly flows were the most significant (loading absolute values > 0.30) variables along with the annual mean number of days with zero flow. The September values had the largest loading (0.3851). On the second PCA axis the mean annual frequencies of low and high flows had the largest loadings (-0.4969 and -0.4215, respectively). On the third PCA axis the mean annual seven-day high and low flows were the most significant variables (0.5265 and 0.5026, respectively).

The variables with the highest loadings were used in a cluster analysis to group the variables and further reduce the number of variables to be used in the CCA with the fish community data. Three groups were evident and corresponded to the groupings from the PCA (fig. 14). The first cluster, which contained the variables important on PCA axis 1, were the coefficients of variation of mean monthly flow and the mean annual duration of zero flow. The second cluster contained the mean annual frequencies of high and low flow, and the third cluster contained the mean annual seven day high and low flows and the mean annual Julian date of low

Table 12. Summary of principal components analysis (PCA) of hydrologic variability.
[Only variables with at least one loading (absolute value) greater than 0.30 are reported; CV, coefficient of variation].

Variable	Principal Component 1 Loadings	Principal Component 2 Loadings	Principal Component 3 Loadings
CV October flow	0.3482	0.2023	-0.1240
CV January flow	.3529	.2544	.0151
CV April flow	.3300	-.0533	-.0459
CV July flow	.3284	-.1740	.1489
CV August flow	.3156	-.2387	.1380
CV September flow	.3851	.1009	-.0019
Seven day high flow	-.1611	.2915	.5265
Seven day low flow	-.1963	.3089	.5026
Mean number of days of 0 flow	.3697	.2241	.0484
Mean jullian date low flow	.0994	-.1526	.3723
Mean jullian date high flow	-.1000	.2244	-.4101
Mean annual frequency of low flow (less than the 25th percentile of period of record)	-.0325	-.4969	.1873
Mean annual frequency of high flow (greater than the 75th percentile of period of record)	.1229	-.4215	.2132
Summary			
Eigenvalue	5.8721	2.6464	1.7350
Proportion of total variance	.3915	.1764	.1157
Cumulative proportion	.3915	.5679	.6836

flow. The mean annual Julian date of high flow did not group well with the other variables.

One variable from each group was selected for analysis with the fish community data. The coefficient of variation of mean April flow was selected from the first cluster group to represent the variability in flow during the spring spawning season. The frequency of low flow was selected from the second cluster group as indicative of the prevalence of low-flow conditions and the mean annual seven day high flow was selected from the third cluster group to indicate high-flow conditions.

Only data from 1992-94 are used in the analysis with the fish data. The mean annual frequency of high flow (the 75th percentile) was added to account for the unusual frequency of high flows during the period preceding fish sampling.

Water Quality

Water-quality data were analyzed in a manner similar to the habitat data. Maximum, minimum, median, and coefficient of variation values were determined for the data (table 2). Maximum and minimum values were selected because effects of stress on biota should be more recognizable from extreme conditions (for example, minimum dissolved oxygen or maximum temperature). Median values were selected for testing as a potential surrogate for maximum and minimum values. The coefficient of variation was used as a measure of the variability of the water-quality conditions. A correlation matrix was constructed to examine the relations between the various constituents. No data transformations were used other than to standardize the data. The correlation matrix of water chemistry and water-quality parameters (maximum, minimum, and median values) indicated high

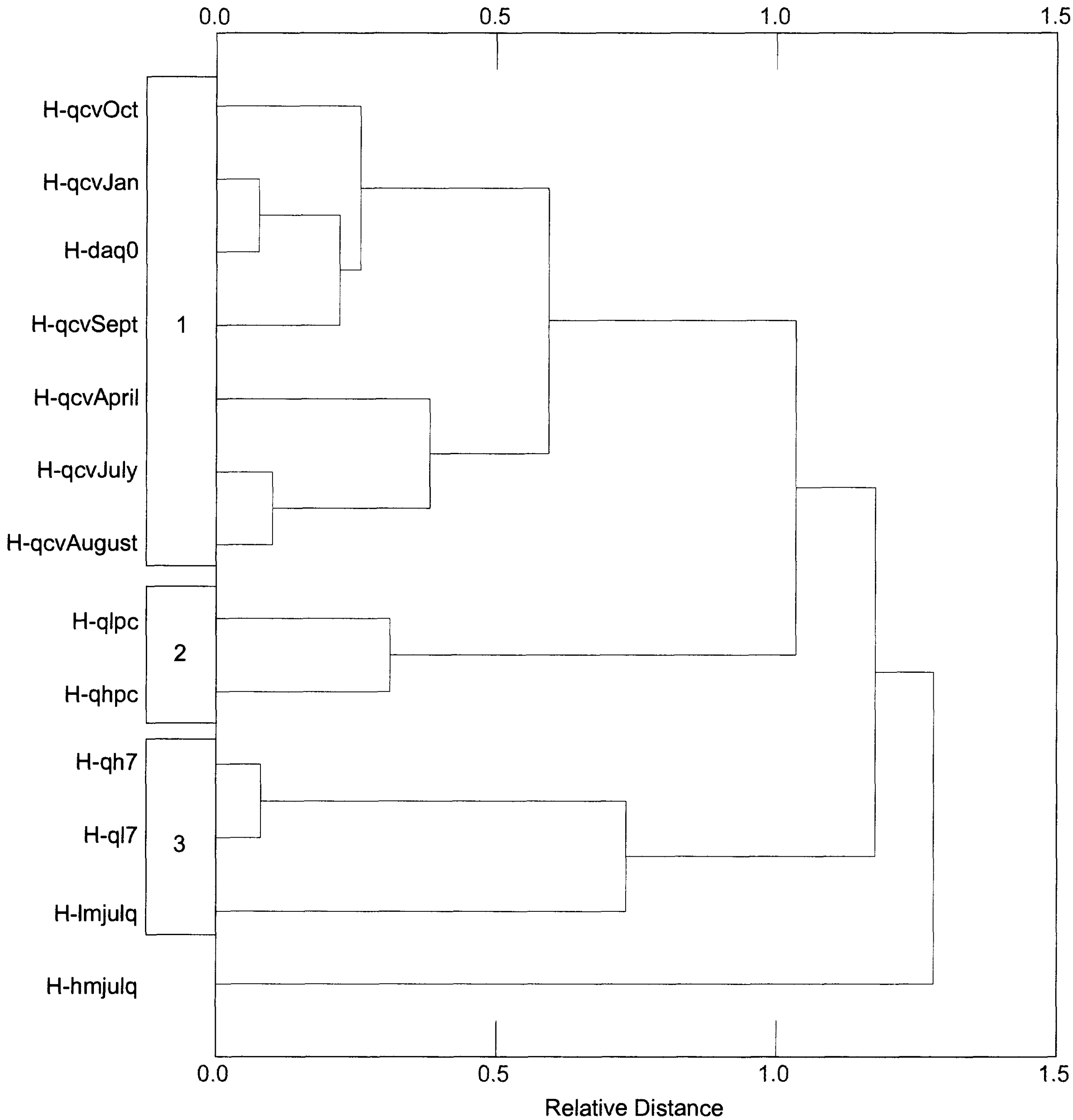

Figure 14. Cluster analysis dendrograph of hydrologic variability loadings from principal components analysis (PCA). [Refer to table 2 for definitions of variable abbreviations.]

correlations ($r>0.6$, $p<0.05$) among certain constituents. The ionic chemistry of the Red River Basin is based on the concentrations of major ions such as calcium, sulfate, and chloride. Specific conductance, which measures ionic strength or the relative amount of ions in solution, correlated well with both calcium and sulfate regardless of the level of the variables used: maximum, minimum, or median. Maximum total phosphorus correlated with maximum total nitrogen. Because of these correlations, a number of variables were deleted for the cluster analysis when surrogate variables were identified through correlation.

Data were not available for all sites where habitat and fish community data were collected. All available data from fourteen sites were entered into a cluster analysis. Three cluster and principal component analyses were performed to determine the best separation of sites and the variables with the largest loadings on the first three PCA axes. Median values were not used even though results using median values were comparable to results using maximum and minimum values. A certain amount of biologically relevant information is lost when using median verses maximum and minimum values. In the first analysis maximum and minimum values were used. Specific conductance was used as a surrogate variable to represent both calcium and sulfate at all levels because of the correlations between specific conductance and these ions. No other reductions in the number of variables were used even though maximum phosphorus correlated well with total nitrogen. Both variables were used to better indicate fertilizer use in this highly agricultural basin. In the second analysis the coefficients of variation of the values were used, and specific conductance was included. The third analysis refined the first by (1) removing the minimum values for water temperature (0 ^{0}C at all sites) and (2) removing specific conductance and including calcium and sulfate to provide more expression of specific anion and cation concentrations.

The first cluster analysis of the water-quality data (minimum and maximum values) separated the rivers into two major groups with three rivers as outliers to both groups. The outliers were the Pembina, Bois de Sioux, and Turtle Rivers. In the Turtle and Bois de Sioux Rivers, water chemistry is greatly influenced by upstream impoundments. The discharge from Lake Traverse on the Bois de Sioux River exhibits high specific conductance as a result of the concentration of sulfate (median=370 mg/L, minimum= 130 mg/L, and maximum=790 mg/L; (L.H. Tornes, USGS, written communication, unpublished data, 1996) which chemically differentiates this site from other sites sampled. The Pembina River carries the highest sediment load of all the tributaries to the Red River (M.E. Brigham and L.H. Tornes, USGS, written communication, unpublished data, 1996). The other two groups separated along geographical boundaries. One group contained sites in Minnesota: the Otter Tail, Roseau, and Red Lake Rivers, all of which drain extensive wetlands or contain flow-through lakes. The other group contained sites in North Dakota and the mainstem of the Red River. The only exceptions were the Snake and Wild Rice Rivers which clustered with the North Dakota and mainstem sites. The PCA indicated moderately high (absolute values <0.38) loadings for variables on principal component Axis 1. Loadings on principal component Axis 2 were not much better (absolute values < 0.39). The strongest loadings on principal component Axis 1 were maximum specific conductance (0.36), maximum ammonia plus organic nitrogen (0.37), and maximum and minimum total phosphorus (0.37 and 0.36, respectively). The largest loadings of the second principal components axis were maximum suspended sediment fines (0.35), minimum dissolved oxygen (0.37), and minimum ammonia plus organic nitrogen (-0.39). The largest loading from the third principal components axis was maximum dissolved oxygen (0.47). The first three principal components accounted for 55 percent of the variation in the data.

In the second analysis the coefficients of variation of all the variables were used. The cluster analysis identified three outliers, the Pembina, Turtle, and Roseau River sites. The remainder of the sites did not cluster well other than the mainstem sites on the Red River. The variability of water quality appears to be different at the three outlier sites than at the other sites. On the first principal components axis, four variables had loadings >0.30: specific conductance (0.40), dissolved calcium (0.35), dissolved chloride (0.35), and pH (0.34). The largest loadings (>0.35) on principal component Axis 2 were dissolved organic carbon (0.48), ammonia plus organic nitrogen (0.43), and dissolved calcium (0.39). Dissolved nitrite plus nitrate had the largest loading (0.52) on the third axis. The first three principal components account for about 63 percent of the variability of the water-quality data.

The third analysis was similar to the first because it used maximum and minimum values. This analysis (fig. 15) identified the same three outliers: the Bois de Sioux, the Pembina, and Turtle River sites. The remainder of the rivers clustered into three groups. The first group contained the Otter Tail and Roseau River sites. The second group contained four mainstem Red River sites,

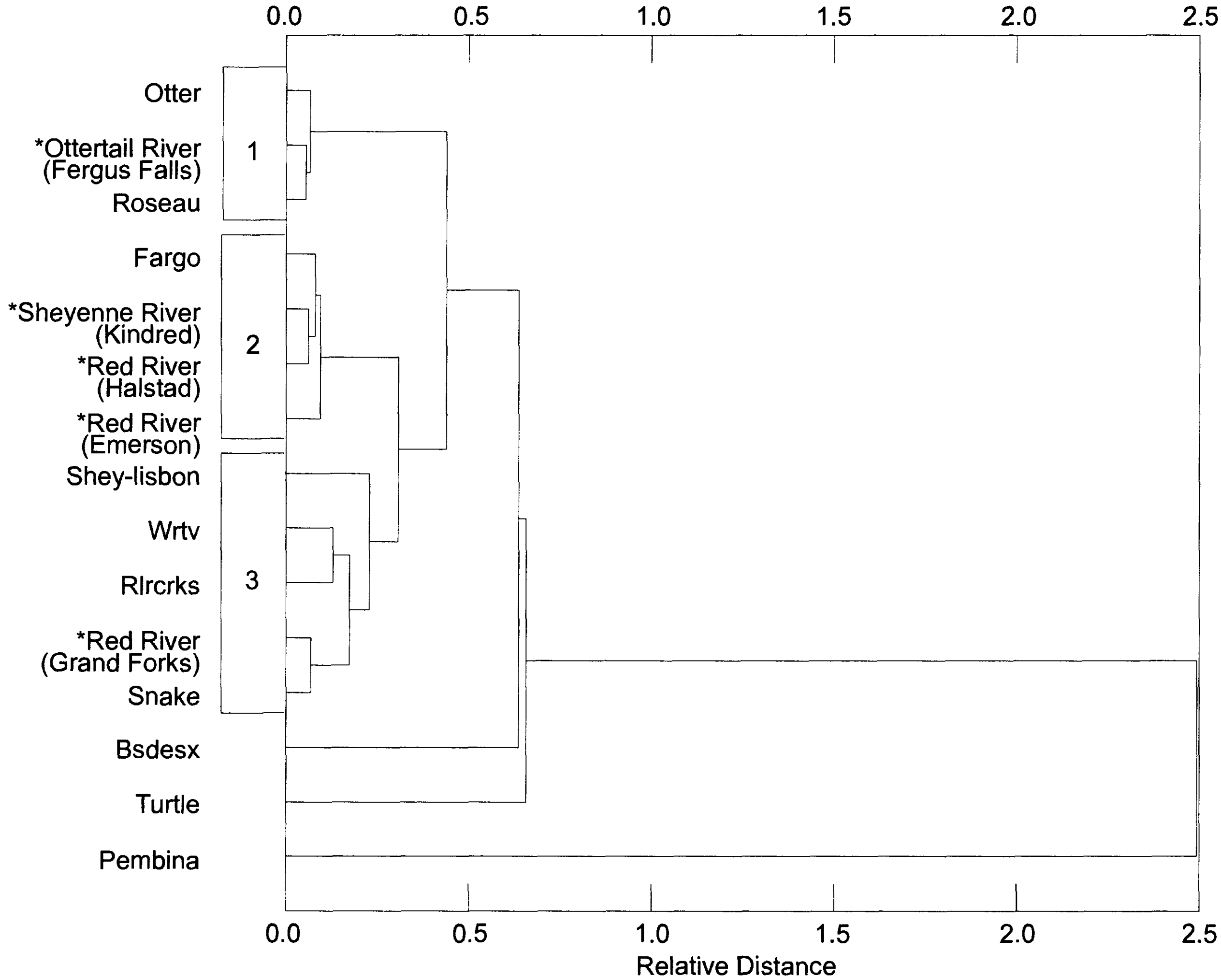

Figure 15. Cluster analysis dendograph of ecological sites grouped by water quality. [Refer to Table 1 for definitions of site abbreviations. *Sites with chemistry data only.]

and the third group contained the all the other sites including a mainstem site, the Red River at Grand Forks, North Dakota. On the first principal components axis the variables with loadings >0.30 (table 13) were the same as in the first analysis except that maximum dissolved calcium and maximum and minimum dissolved sulfate replaced specific conductance and have loadings of 0.3324, 0.3545, and 0.3191, respectively. The other high loadings result from maximum and minimum total phosphorus (0.3560 and 0.3306, respectively). Maximum ammonia plus organic nitrogen had a loading of 0.3272. The second principal components axis had minimum dissolved organic carbon and minimum ammonia plus organic nitrogen with the highest loadings (-0.3606 and -0.4136, respectively). The highest loading from the third principal components axis was maximum dissolved oxygen (0.4660). These three principal components explained about 58 percent of the variability in the data. The classification of the rivers into groups was based on the principal components from the third analysis, which were able to discriminate among sites as well as the other analyses, but the third analysis produced groupings of the sites that corresponded more closely to the cluster groups.

The variables with the highest loadings were used to cluster the variables and reduce the number of variables used in the CCA with the fish community (fig. 16) A single variable from each cluster group was chosen to represent the group of variables within the cluster. Unlike the instream habitat data, there did not appear to be a single underlying characteristic or logic for each of the groups. One group contained minimum dissolved oxygen and minimum dissolved calcium. Another group contained all the other variables except three which grouped together (minimum dissolved organic carbon,

Table 13. Summary of principal components analysis (PCA) of water-quality variables.
[Only variables with at least one loading (absolute value) greater than 0.30 are reported].

Water-quality variable	Principal component 1 loadings	Principal component 2 loadings	Principal component 3 loadings
Total phosphorus-maximum	0.3560	-0.0078	-0.0793
Dissolved sulfate-maximum	.3545	.0733	.0200
Dissolved calcium-maximum	.3324	-.0433	-.0776
Total ammonia and organic nitrogen-maximum	.3272	-.1191	-.1762
Total phosphorus-minimum	.3306	.0142	.1077
Dissolved sulfate-minimum	.3191	.2468	.0997
Total ammonia and organic nitrogen-minimum	.1141	-.4136	-.0081
Dissolved organic carbon-minimum	.0541	-.3606	-.2739
Fine particulates-minimum	.0699	-.3080	.2504
Dissolved calcium-minimum	.0128	.3396	.0477
Dissolved oxygen-minimum	.0357	.3526	-.2848
Dissolved oxygen-maximum	-.0363	.0538	.4660
Dissolved chloride-maximum	.2399	-.0277	.3498
Dissolved chloride-minimum	.2235	.2343	.3293
Summary			
Eigenvalue	6.5177	4.3877	2.5786
Proportion of total variance	.2834	.1908	.1121
Cumulative proportion	.2834	.4741	.5862

minimum percent of particulates smaller than sand, and minimum total nitrogen from ammonia plus organic nitrogen). Maximum dissolved oxygen was an outlier.

The variables selected to be the representatives of each of the groups were minimum dissolved oxygen, maximum total phosphorus, and the minimum percent of particulates smaller than sand. All three variables have biological implications for fish. Minimum dissolved oxygen concentrations can cause severe stress or mortality, maximum total phosphorus concentrations relate to fertilizer use and primary productivity of the rivers, while the percent of particulate sediment smaller than sand indicates the potential for siltation effects.

Fish

The cluster analysis of the Jaccard coefficients of similarity produced four major clusters that contained most of the reaches sampled (fig. 17). The first cluster contained the streams with the greatest number of species regardless of location or stream size. The second cluster contained four sites in the east-central portion of the basin. All four sites are either in, or border on, the North Central Hardwood Forests ecoregion. Both Wild Rice River sites are within the ecoregion while the Sandhill River site and the Clearwater River at Berner site are just adjacent to the ecoregion (fig. 4). The third cluster contained all the other sites in North Dakota that were not in the first cluster group. Additionally, there were two Minnesota sites closely related to this third group, the Clearwater River at Red Lake Falls and the Two Rivers site. These two sites tended to contain fewer species (12 and 9, respectively) than either the Minnesota or North Dakota sites in the second cluster group (means of approximately 15); however, the species were similar, resulting in the relatively close grouping. The third major cluster group contained the mainstem Red River sites, including the Bois de Sioux River site. The remaining three sites did not group well with any other sites. The Snake River had the fewest species of any of the sites, but is located in the central part of the basin. The Roseau and Otter Tail River sites

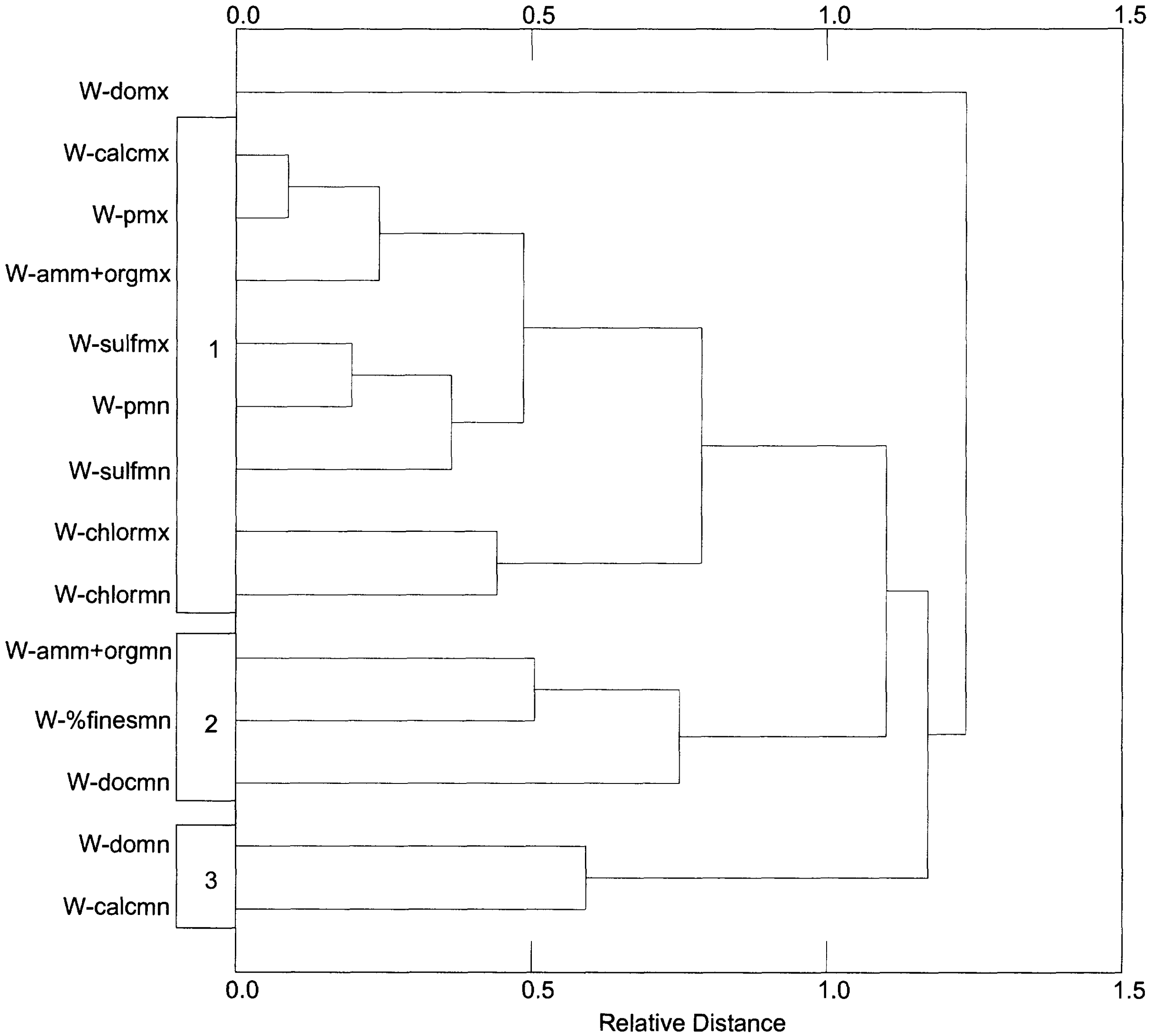

Figure 16. Cluster analysis dendrograph of water-quality variable loadings from principal components analysis (PCA). [Refer to table 2 for definitions of variable abbreviations.]

are in different ecoregions from most of the other sites. The Roseau River is in the Northern Minnesota Wetlands ecoregion and the Otter Tail River site is in the Northern Lakes and Forests ecoregion. The dissimilarity of species collected from rivers in Minnesota, North Dakota, and the mainstem of the Red River partially accounts for the cluster analysis results.

The clustering of the multiple reach sites indicates that sampling was consistent. The results of the fish community clusters in conjunction with the instream habitat clusters indicates that, for the most part, the multiple reaches were similar in terms of the physical habitat and the fish community samples collected at each reach.

A second procedure was applied to the fish data. DCA was used to distinguish the sites by both species composition and abundance of individuals per species (Ter Braak, 1987). The fish data were edited to remove any species which occurred at less than 5 percent of the sites (Gauch, 1982) This procedure yielded site distinction similar to the cluster analysis (fig. 18). The dominant feature of the analysis is the distinction of sites along gradients. The similarity of the results was due to the distinction of the sites by river size and species associations. Both the cluster analysis and the

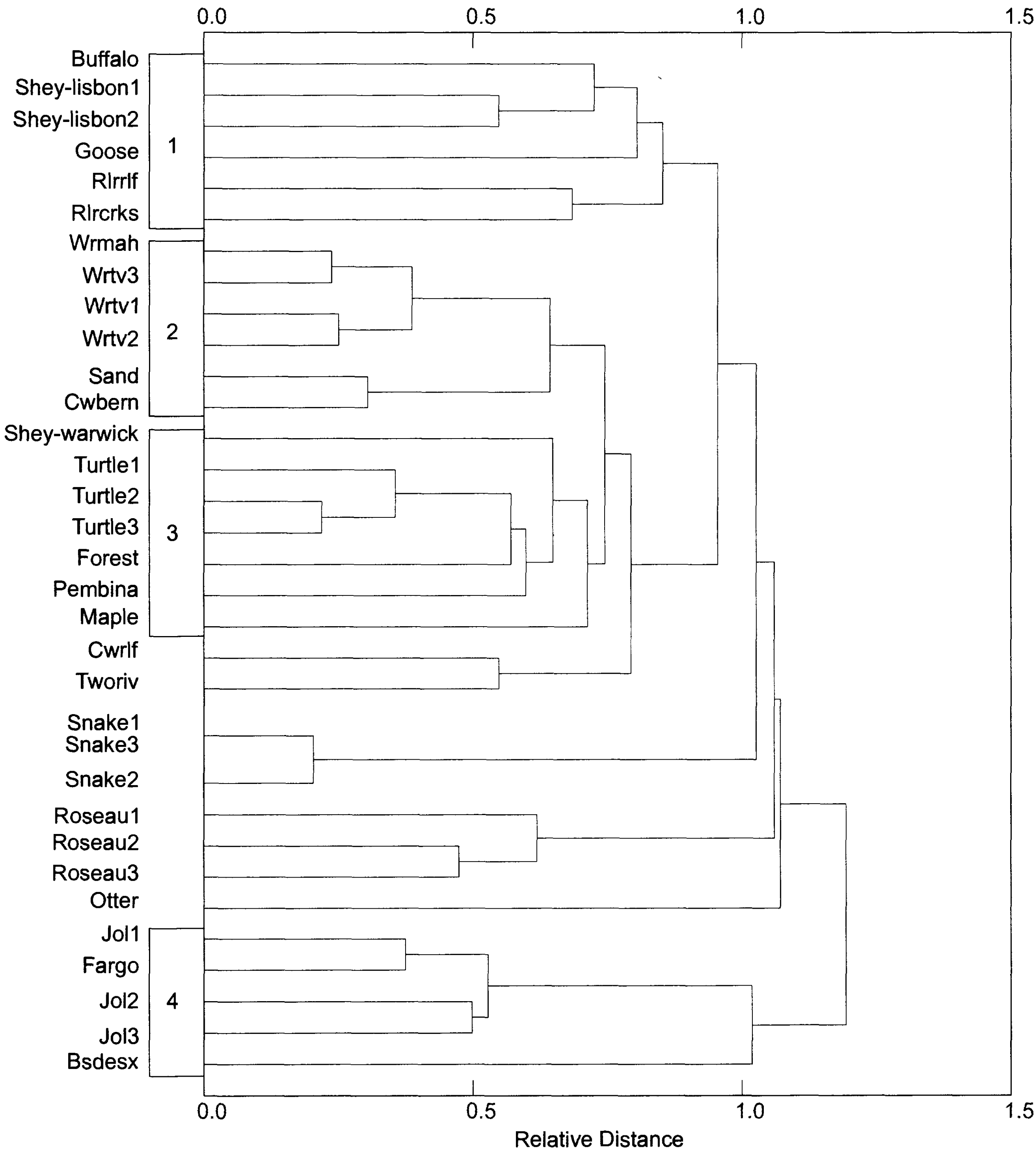

Figure 17. Cluster analysis dendrograph of fish species similarity at ecological sampling sites in the Red River of the North Basin. [Refer to table 1 for definitions of site abbreviations.]

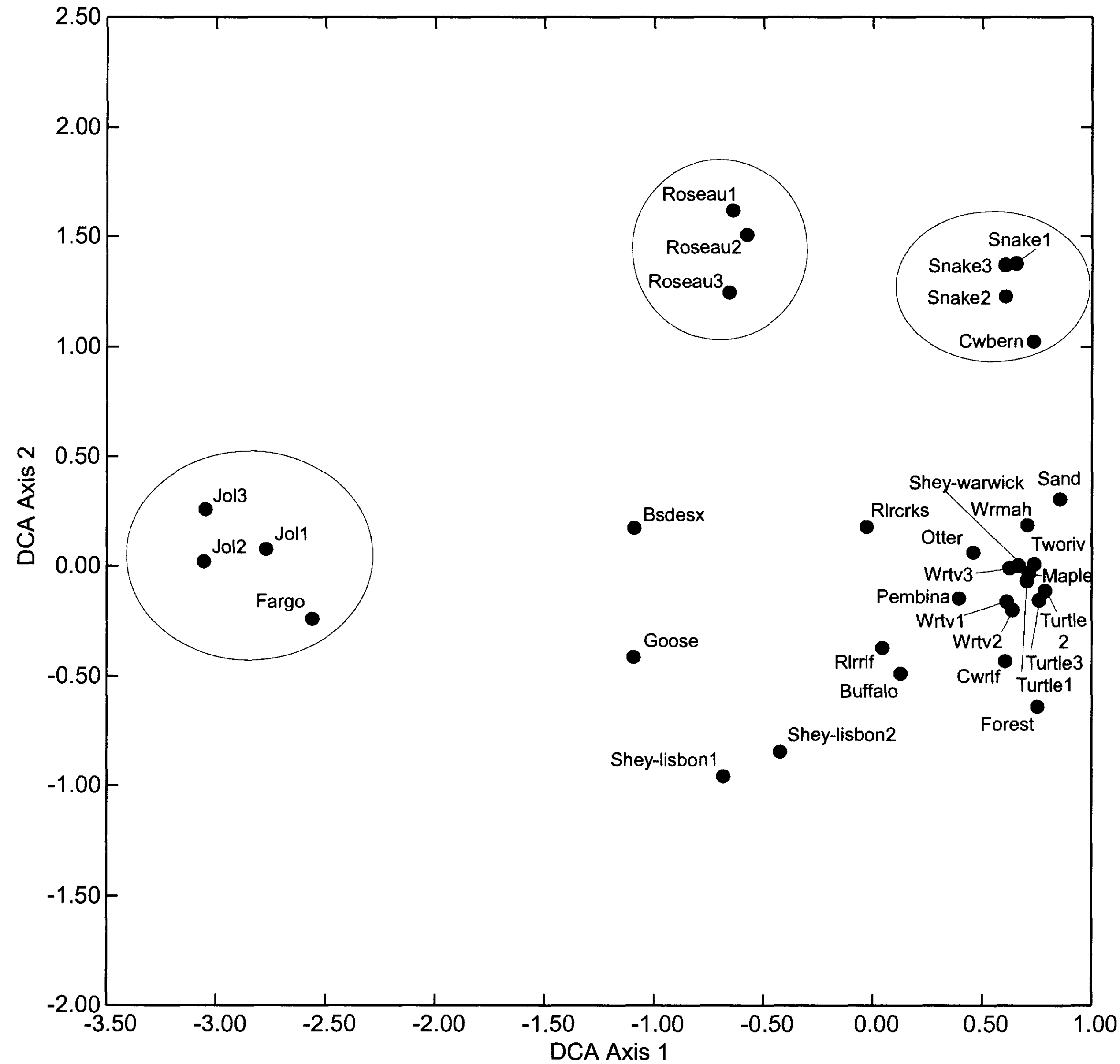

Figure 18. Location of ecological sites on detrended correspondence analysis (DCA) axes. [Refer to table 1 for definitions of site abbreviations.]

DCA grouped the mainstem Red River sites together (the three reaches at Joliette and the Fargo site). The three Roseau River reaches grouped together in both analyses as did the three Snake River reaches. The Clearwater River at Berner site grouped with the Snake River. The DCA and cluster analysis seem to be in agreement regarding site distinction and classification.

The reach scores from the DCA were use to confirm environmental variable selection. DCA scores from the first DCA axis were compared to each environmental variable in a correlation matrix to confirm that variables identified as important in site classification by cluster analysis and PCA were also important with respect to the fish community. The amount of variation in the fish community explained by the ordination of the fish community data in the DCA was about 25 percent. Therefore, correlation values which were >0.50 ($p<0.01$) were considered significant.

Within the instream habitat data set, the only significant correlations were between the fish community DCA scores and two measures of river size: mean channel width and mean depth. Both of these measures of river size correlated well with the DCA fish

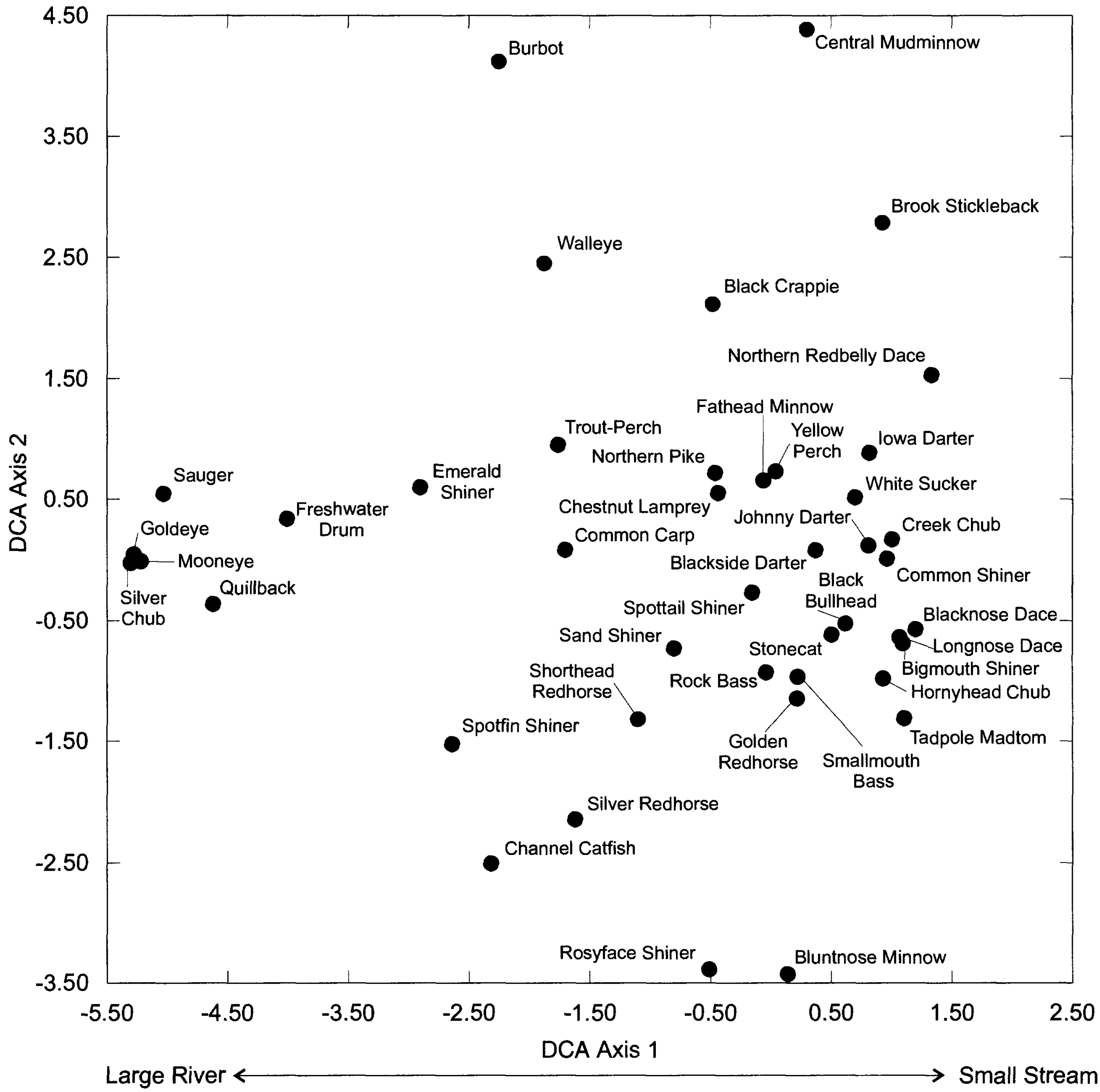

Figure 19. Species distribution with stream size as determined from detrended correspondence analysis (DCA).

community scores. The correlation coefficients were 0.77 ($p<0.001$) for mean channel width and 0.84 ($p<0.001$) for mean depth.

Correlations between fish community DCA scores and segment and basin characteristics identified some of the same characteristics that were identified by the cluster analysis and PCA. Riparian tree density on both banks correlated with DCA scores (right bank $r=0.71, p<0.001$; left bank $r=0.54$, $p<0.007$). Right bank riparian zone discontinuity was negatively correlated with fish community DCA scores ($r=-0.57$, $p<0.005$). The percent of land as riverine wetlands, which was selected as a representative variable based on the cluster analysis and PCA, correlated with fish community DCA scores as well ($r=0.60$, $p<0.003$). One basin characteristic which correlated well with fish community DCA scores was the total length of ditches in the basin upstream of the site ($r=0.58$, $p<0.004$).

The correlation of the water-quality data and fish community DCA scores identified almost the same variables as the cluster analysis and PCA. Both analyses identified minimum dissolved oxygen as important. The correlation with the fish community DCA scores was r=-0.51 ($p<0.01$). Both analyses also identified the minimum percent of fine particulates smaller than sand ($r=0.59$, $p<0.004$) with the DCA scores. The only other water-quality variable that produced a significant correlation was minimum specific conductance, which correlated negatively with fish community DCA scores ($r=-0.70$, $p<0.003$).

All the hydrology statistics that either increase or decrease with river size were well correlated with fish community DCA scores ($r>0.76$, $p<0.001$). Those statistics include any measures of high, low, or mean flows.

The fish species of the Red River Basin may be grouped by their relative position along the various gradients determined for instream and terrestrial habitat, water quality, and hydrologic variability (fig. 19). Large river species include sauger, quillback, mooneye, goldeye, silver chub, freshwater drum, and emerald shiner. Small stream species are longnose and blacknose dace, johnny and Iowa darter, tadpole madtom, common and bigmouth shiner, hornyhead and creek chub, brook stickleback, and northern redbelly dace. The remaining species appear to grade between small, intermediate, and large streams.