



Project Summary

A Computerized System for the Evaluation of Aquatic Habitats Based on Environmental Requirements and Pollution Tolerance Associations of Resident Organisms

Clyde L. Dawson and Ronald A. Hellenthal

The Environmental Requirements and Pollution Tolerance (ERAPT) system is a computerized retrieval and analysis system for environmental information on aquatic organisms. It can be used to predict organism assemblages based on environmental conditions, to describe environmental characteristics associated with plants and animals which inhabit specific sites, and to compare site characterizations with associated biological communities for inconsistencies resulting from organism misidentification or erroneous habitat information. It can also identify species that are particularly sensitive to specific sets of environmental conditions and can predict changes in species composition that are likely to result from environmental disturbance.

The system has been developed for an IBM 370/3033* computer system. It can be used interactively to predict and evaluate aquatic habitats and organism assemblages. ERAPT computerized data bases contain 57,345 biological, physical, chemical, and distributional characteristics drawn from 883 sources for 1,691 species of North American diatoms, blue-green algae, the insect orders Ephemeroptera (mayflies) and Plecoptera (stoneflies), the

dipteran family Chironomidae (midges), and the fishes of U.S. Environmental Protection Agency Region 5 (North Central U.S.).

The system was tested by comparing its prediction of environmental conditions based on the dominant fish community with environmental characteristics obtained from aquatic habitat surveys for 64 sites on three river systems in Ohio. An overall successful prediction rate of 95% was achieved. Where disagreements between predicted and reported environmental characterizations occurred, they almost always were among very similar tolerance ranges or habitat types.

This Project Summary was developed by EPA's Environmental Research Laboratory, Corvallis, OR, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

Aquatic organisms long have been recognized as potentially useful indicators of habitat conditions and water quality. This is due to their ability to reflect conditions through time, to demonstrate the effects of disturbances after the environment has returned to apparently normal physical and chemical conditions, to

*Mention of trademarks or commercial products does not constitute endorsement or recommendation for use.

integrate the effects of many different environmental factors and their interactions simultaneously, and to provide a living context for considerations of environmental quality. The basic requirements for indicator organisms are that they must be sensitive to environmental changes and have life spans and generation lengths which are appropriate for use in environmental assessment, long enough to reflect intermittent or occasional disturbances, and short enough to subject sensitive life stages to adverse environmental conditions.

Although organisms have been a fundamental component of aquatic habitat surveys, they play a relatively minor role in aquatic assessment. One reason may be the difficulties encountered in comparing and summarizing information obtained by different researchers. This may be due to variability in environmental requirements and habitat associations for some species making them unsuitable as environmental indicators or the result of inconsistent methods of data collection or analysis. A second factor may be the diverse outlets for publication of these data; often appearing only in progress and summary reports receiving little dissemination. Another factor is the difficulty in developing quality control procedures for organism identifications. This has resulted in great variation in the quality and reliability of published associations between organisms and specific environmental conditions.

Environmental data and reports vary widely in quality and reliability. Standardized methods for data acquisition and analysis and quality control procedures are reasonably well established for most biological, chemical, and physical environmental data. However, the association of specific organisms with these parameters rests on the reliability of species identifications performed by investigators of varying backgrounds, experience, and expertise. At present, the primary quality assurance mechanism for organism identifications is the competence and dedication of individual researchers. While independent verification of voucher specimens by taxonomic experts is possible, this is a procedure which, even on a small scale, is time consuming and costly and, on a large scale, would significantly exceed our present resource of taxonomic experts.

Publications providing summary information on the environmental relationships and tolerances of aquatic organisms also may be useful in identifying potential indicator organisms and in establishing

tolerance levels of organisms to specific environmental factors. However, unless the reference publication survey is extremely exhaustive or very selective, it may be difficult to establish useful trends among the vast amounts of often contradictory information reviewed. Furthermore, the fixed format of printed reports severely limits the ways in which their data can be used. It is often difficult to determine which environmental requirements are consistent within a genus or higher taxonomic level and which environmental inconsistencies at the specific level could be due to errors in data collection or identification.

A computerized storage, retrieval and evaluation system for these data has been developed to enhance their usefulness by allowing queries based on taxonomic association, environmental parameters, or a combination of these factors. Since the environmental information for diverse taxonomic groups can be standardized, it is possible to answer questions concerning biological communities and their environments. A major application of this system is the development of lists of taxa associated with specific environmental conditions which can serve as reference communities for environmental scientists. This system also can be used as a device for scrutinizing environmental data by comparing the known tolerances of organisms reported in monitoring and impact studies with the physical and chemical characteristics of the habitats from which they were collected. Discrepancies encountered in these comparisons are flagged as potential errors in data collection or specimen identification. Another quality control device possible with these data is comparing the known ecological parameters associated with combinations of taxa reported to be collected together at individual sites. Inconsistencies in the environmental requirements of these taxa also can be flagged as potential errors in identification.

The Environmental Requirements and Pollution Tolerance (ERAPT) information retrieval and analysis system permits computerized prediction and evaluation of habitat characteristics and organism assemblages. ERAPT system data bases contain 57,660 parameter values representing environmental characterizations of 402 genera and 1,694 species of aquatic organisms in the diatoms, blue-green algae, the insect order Ephemeroptera and Plecoptera, and the dipteran family Chironomidae, and the fish species of U.S. Environmental Protection Agency (EPA) Region 5 (North Central U.S.). The en-

vironmental characterizations used by the ERAPT system have been derived from 883 data sources, primarily from the published literature (Table 1). This represents a mean of 34 environmental parameters from six data sources per species, with four species per genus. Information for the environmental characterization of diatoms, bluegreen algae, chironomids, Ephemeroptera and Plecoptera were obtained from the EPA Environmental Monitoring and Support Laboratory, Cincinnati, Ohio. Characterization of Region 5 fishes from primary and secondary literature sources was done as part of this project.

System Operation

The ERAPT system consists of an integrated set of computer programs for the storage, retrieval, and manipulation of environmental requirements and pollution tolerance information on aquatic organisms. It has been developed for use on an IBM 370/3033 computer system at the University of Notre Dame. Data are stored and manipulated as hierarchically related environmental requirements and pollution tolerance categories representing tolerance ranges to specific pollutants or environmental conditions, geographic locations, general or specific habitat characteristics, and periods of appearance, emergence, or greatest abundance. At present the system uses 21 heading categories (stage, abundance, pollution tolerance, optimal growth period, reproductive season, reproductive behavior, feeding behavior, hardness, salinity, nutrients, degradable organics, pH, oxygen, temperature, turbidity, current, general habitat, specific habitat, bottom type, ecosystem regions, political geography) divided into 303 specific parameters. These categories and parameters were adapted from a set developed by the Aquatic Biology Section, EPA Environmental Monitoring and Support Laboratory (Cincinnati) for use with macroinvertebrates and diatoms, and were expanded for fish and other aquatic organisms as part of this project in collaboration with the staff of the EPA Corvallis Environmental Research Laboratory.

Data for ERAPT are encoded on tabular forms which can be produced by the system. These forms contain boxes corresponding to specific parameters such as stage, feeding behavior, or tolerance to environmental conditions for an individual species from a single reference source. The ERAPT system reads the tabular forms as digitized X-Y coordinate values corresponding to each mark on a form.

Table 1. Summary totals for information in the ERAPT data bases.
 BLGR, Bluegreen algae; DIAT, Diatoms; CHIR, Chironomids;
 EPHE, Ephemeroptera; PLEC, Plecoptera; FISH, Fishes of EPA Region 5

	BLGR	DIAT	CHIR	EPHE	PLEC	FISH	Total
Authors	420	48	34	200	125	56	883
Records	221	342	306	459	450	446	2,224
Genera	50	50	74	60	86	82	402
Species	161	295	232	399	364	243	1,694
Citations	2,289	2,726	529	1,207	1,767	1,000	9,518
Entries	4,364	5,510	8,156	6,806	7,685	25,139	57,660

The digitized data are standardized for the ERAPT system. During this process both direct and hierarchical correspondences between categories are established. For example, one set of data entry forms may contain the environmental category Mesosaprobic, whereas another may divide this category into alpha and beta ranges. During this step these parameters are associated so that a parameter shown at a lower level of the environmental category hierarchy is included at all upper levels. For example, the heading salinity is divided into 11 categories with the categories Mesohalobous and Oligohalobous further subdivided into two and three parameters, respectively (Figure 1). This structure permits the ERAPT system to store, manipulate and use environmental information differing in precision without sacrificing the most reliable data.

The next procedure is the creation of a searchable data base. At this point the various components of the system are linked. The taxonomic categories are hierarchically connected in a matter similar to that described previously for environ-

mental headings and parameters. This enable queries at any level of the taxonomic hierarchy. The linking operation involves the association of the taxonomic names, environmental headings and parameters plus their definitions, author citations and references, environmental data, and a series of 2 to 8-character abbreviations which are used in the query and information retrieval process. The lowest taxonomic level directly accessible by the system is subspecies. However, data from individual sources are also available.

The system uses two summaries of the information contained in reference sources for each taxon, a logical sum and logical product. The logical sum includes all environmental parameters given by any source for each taxon in the system. For example, if one investigator reported that a species occurs in streams and another reported that the same species occurs in both lakes and streams, the logical sum for general habitat category for that species would include both lakes and streams. The logical product includes only those parameters indicated by all sources containing

environmental requirements or pollution tolerance information within each heading category for a taxon. Therefore, each product summary consists of those environmental parameters which are consistently associated with a taxon in the system. For the general habitat example given above, the product summary would show only lakes, since it was the only parameter indicated by all sources of information for the taxon. Inconsistencies among investigators reporting information about any heading category for a taxon may be obtained by calculating the logical difference (exclusive OR) between the sum and product summaries. In the general habitat example, streams would be considered an inconsistent parameter since one source used the parameter and the other did not. For taxonomic categories above species, similar data summaries are maintained based on logical sums of the information for all lower taxonomic levels. The system also maintains the number of data sources which were summarized for each taxon and environmental category.

Information retrieval and analysis are accomplished through an interactive program which uses simple commands for queries of the data base. These commands may be used to predict groups of organisms having a select set of environmental characteristics. Species may be evaluated for consistency of associated environmental information, providing lists of potential indicators. Groups of organisms also may be defined using abbreviated taxonomic names for characterization and evaluation. Species or groups of species may be compared to determine the similarities and differences in their environmental requirements. Information may be selected by taxonomic level, number of sources, or groups of environmental parameters.

The program provides information on program commands, definitions of environmental retrieval codes and a profile of the searches performed during the session. Individual search statements or complete terminal sessions may be saved and later recalled for use with data bases containing information on other groups of organisms. The program can retain an almost unlimited number of saved searches and can provide directories and lists of these upon request. The program will permit customization for a variety of printing and nonprinting computer terminals and can be directed to pause between output pages. Input may come from sources other than a terminal to permit other programs to interface directly with

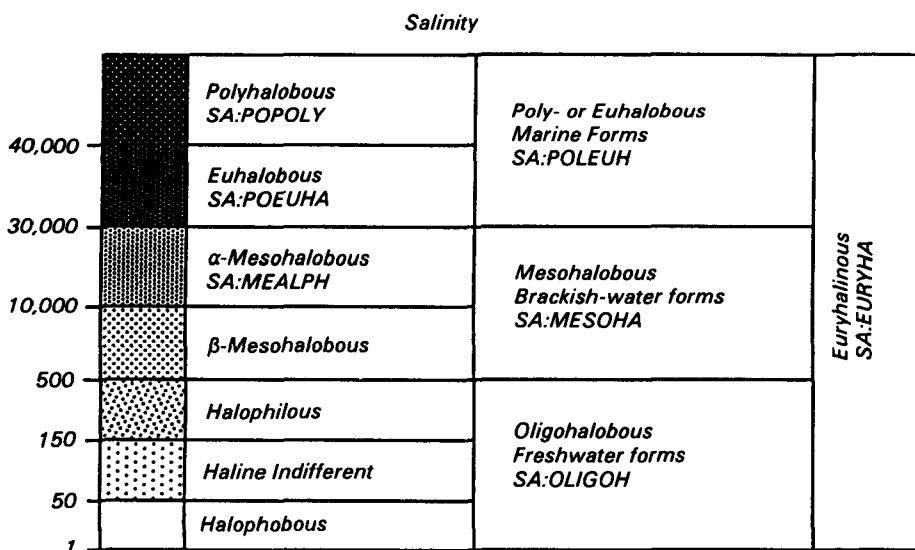


Figure 1. Hierarchy diagram for salinity in mg/l.

the system. Output may also be redirected to files or to a high-speed lineprinter.

Validation Methods

Survey data for verification of the ERAPT system were obtained from the State of Ohio Environmental Protection Agency for 64 sites from 20 rivers, streams, and tributaries in the Cross Creek and Yellow Creek, Blanchard River, and Tuscarawas River systems in southeastern and northwestern Ohio. For each site a list of fish species and abundances and descriptive information on the habitat was provided. This included the dissolved oxygen, temperature, pH, turbidity, bottom type, and general and specific habitat characteristics of most sites, with current, degradable organics, and pollution information provided for some of the sites.

The five numerically dominant species at each site were entered into the ERAPT system and used to predict local environmental conditions. These species accounted for at least one half of the total fishes collected and, therefore, were taken to be representative of the environmental conditions present within the habitat sampled. The ERAPT system's compare function was used to identify the environmental characteristics common to each fish community. This information was used for comparison with the environmental description of each habitat.

Habitat characterization information for each collecting site was transformed into ERAPT environmental parameter codes directly from field survey forms. These data then were compared with the environmental descriptions predicted by the system from the associated dominant fish community for each category of environmental characteristic. The compared data were considered to be in agreement if they shared common environmental parameters. For example, a general habitat predicted by the fish assemblage as lake or large river was considered to be in agreement with a habitat characterized as either large river or lake. Likewise, a community predicted bottom type of gravel was considered to be in agreement with a habitat characterized as gravel, sand, and silt.

Validation Results

In comparing 300 environmental characteristics obtained from field survey forms to those predicted by the ERAPT system from the dominant fish community, 285 showed agreement for an overall successful prediction rate of 95% (Table 2). The successful prediction rates for the Cross Creek and Yellow Creek, Blanchard

River, and Tuscarawas River systems were 95%, 97%, and 93%, respectively. There was 100% agreement between predicted and observed characterizations for temperature, current, and degradable organics in all three river systems. Those environmental parameters not in complete agreement between observed and predicted characterizations were: general habitat (98%), dissolved oxygen (96%), specific habitat and bottom type (95%), and pH and turbidity (88%).

While there were some disagreements between predicted and reported environmental characterizations, they almost always were among very similar tolerance ranges or habitat types. Most of these disagreements were considered to be minor by the staff of the State of Ohio Environmental Protection Agency considering the precision of the original measurements. In addition, only one site had discrepancies between the observed and predicted characterizations of more than one kind of environmental parameter. This was site number 1 of Yellow Creek where there were disagreements in both specific habitat and bottom type. Overall, ERAPT system predictions of habitat characteristics from the dominant fish assemblages closely approximated the information obtained from field surveys at all of the sites evaluated.

Future Development

Use of environmental indicator organisms on a regional basis appears to be particularly promising. Assessment systems based on a sound knowledge of a local fauna would seem to have a great probability for success. Development of local or regional environmental data bases is within the capabilities of the ERAPT system. This will be greatly facilitated by development of microprocessor-based versions of the retrieval and analysis programs. Microprocessor implementation of the ERAPT system on computers such as the IBM PC and the Apple Macintosh is feasible and a prototype microcomputer version of the retrieval and analysis program currently is being developed.

An interesting capability of the ERAPT system is the potential to screen environmental data. Detection of apparent inconsistencies among the environmental tolerances of different organisms said to be collected together at a given site can be used both to flag errors in either data collection or organism identifications and to evaluate the reliability of environmental data bases. Taxonomic experts could confirm the identifications of organisms

flagged as inconsistent, providing a cost-effective means of evaluating the quality of biological information. When new environmental associations of organisms are found, the data bases can be updated.

Since the ERAPT system is designed to accept input from files, it could be interfaced with ongoing environmental data collection projects to summarize and evaluate information and to scan for probable changes in environmental conditions. Interfacing programs could convert numerical environmental data and codes unique to local or regional data collection projects so that they can be used with the ERAPT system. Output from the ERAPT system also might be used by other programs for environmental decision making and evaluation. Sources of environmental information within the data bases also can be evaluated. It is possible to identify and eliminate sources that provide environmental information which frequently contradicts that supplied by other workers for the same organisms.

Because the ERAPT programs are independent of the kinds of data and organisms which they can process, the system has many applications in other areas of biological and environmental research. Alternate sets of biological categories or environmental parameters can be developed easily for terrestrial or marine environments or for microbial organisms or higher vertebrate groups. The system also has direct application to behavioral, physiological, and genetic organism information. Because of the structure and internal organization of the data used by the system, there is no limit to the amount of information which can be accumulated or queried for each kind of organism contained within a data base. System capacities are related only to the number of kinds of organisms and data categories which can be accessed and evaluated as a group, and are easily modified within programs. There is also no limit to the number of groups which can be developed. Programs permit transfer of search statements between organism groups if they have compatible data.

Table 2. Results of ERAPT system validation study

River Systems	Cross & Yellow		Tuscarawas		Blanchard		Total	
	No. of Sites	Agreement	No. of Sites	Agreement	No. of Sites	Agreement	No. of Sites	Agreement
DO ₂ mg/l	13	100%	5	100%	6	83%	24	96%
Temperature	18	100%	6	100%	7	100%	31	100%
pH	16	81%	4	100%	5	100%	25	88%
Current	2	100%	3	100%	—	—	5	100%
Turbidity	22	82%	2	100%	8	100%	32	88%
General Habitat	34	100%	11	91%	17	100%	62	98%
Bottom Type	33	97%	10	80%	16	100%	59	95%
Organics	—	—	3	100%	—	—	3	100%
Specific Habitat	31	97%	10	90%	16	94%	57	95%
Pollution	—	—	2	100%	—	—	2	100%

R. A. Hellenthal and C. L. Dawson are with the University of Notre Dame, Notre Dame, IN 46556.

Phil Larsen is the EPA Project Officer (see below).

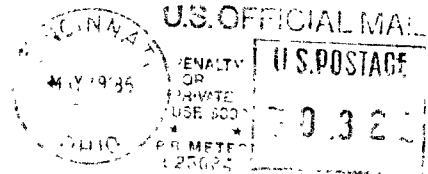
The complete report, entitled "A Computerized System for the Evaluation of Aquatic Habitats Based on Environmental Requirements and Pollution Tolerance Associations of Resident Organisms," (Order No. PB 86-167 343/AS; Cost: \$16.95, subject to change) will be available only from:

*National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650*

*The EPA Project Officer can be contacted at:
Environmental Research Laboratory
U.S. Environmental Protection Agency
Corvallis, OR 97333*

United States
Environmental Protection
Agency

Center for Environmental Research
Information
Cincinnati OH 45268



Official Business
Penalty for Private Use \$300

EPA/600/S3-86/019

0000329 PS
U S ENVIR PROTECTION AGENCY
REGION 5 LIBRARY
230 S DEARBORN STREET
CHICAGO IL 60604