A DATA BASE MANAGEMENT SYSTEM FOR RESEARCH IN POND DYNAMICS

Kevin D. Hopkins, James E. Lannan and James R Bowman

Pond Dynamics/Aquaculture Collaborative Research Support Program, Office of International
Research and Development, Oregon State University, Corvallis, Oregon 97331
USA

ABSTRACT

This paper describes a data base management system used in the analysis and synthesis of data from a global experiment investigating dynamics of tropical farm ponds. Data synthesis is directed toward developing comprehensive farm pond management models. The experimental plan includes standardized data collection on physical, chemical and biological variables on 12 or more ponds at each of seven locations in Africa, Central America, and Southeast Asia.

Field data from each location are compiled on microcomputers at the project sites. Project staff at each site perform routine statistical analyses and presentations of these data sets using statistical and graphics software packages of their choice. Field data are forwarded to a central office for filing on a mainframe computer. More sophisticated data analyses and syntheses (modeling) are done on mainframe computers.

The standardized experiment design and centralized data base management system facilitates analyses of observations within ponds, between ponds within locations, and between locations. This design allows identification analysis of both general and site specific considerations over a broad range of environmental conditions. We propose that aquaculture scientists standardize data collection and management to permit direct comparisons of observations among and between research projects. The resulting comprehensive data base will improve understanding of the dynamic processes that regulate productivity in tropical farm ponds.

INTRODUCTION

The Collaborative Research Support Program (CRSP) in Pond Dynamics/ Aquaculture was established in 1982 with the technical goal of quantitatively describing the physical, chemical, and biological principles of pond culture systems. Seven USA universities participate in this CRSP which is funded by USAID. CRSP research ponds are located at seven stations in six countries: Honduras, Indonesia, Panama, Philippines, Rwanda and Thailand. Each location has at least 12 ponds.

The most recent work plan calls for monitoring 96 variables on several different schedules (Table 1). There are two five-month experiments conducted each year during which each station is expected to collect at least 90,000 observations. If this data collection rate is maintained, over three million observations will be collected during the life of the CRSP.

DAILY	WEELKY AND TWICE WEEKLY	TWICE MONTHLY DIURNAL	OCCASIONALLY
Solar radiation Rain Mean Windspeed Air Temp- max. & min. Evaporation Rate Pond Depth Water Inflow Duration Water Inflow Rate Pond Overflow? Nutrient Input Organic Fertilizer Inorganic Fertilizer Feed No. Dead Fish/Shrimp by species	Oxygen: top, mid, bottom Pond Temp: top, bottom Max Temp: top, bottom Alkalinity Total Hardness pH Kjeldahl Nitrogen Total NH3 and NH4 Nitrate Nitrogen Total Phosphorus Orthophosphate Secchi Disk Visibility Chlorophyll a, b, c Salinity	Time of Day Oxygen: top, mid, bottom Pond Temp: top, bottom pH MONTHLY Primary Productivity B-G Algae Abundance Green Algae Abundance Diatom Abundance Other Algae Abundance Rotifera Abundance Cladocera Abundance Copepoda Abundance Other Zooplankton Abundance Mollusca Abundance Insecta Abundance Decapoda Abundance Other Benthic Organisms Fish/Shrimp Mean Length Fish/Shrimp Mean Weight	WATER: chloride, sulfate, boron, calcium, copper, iron, magnesium, potassium, sodium, zinc FISH/SHRIMP: stocking, harvest, health NUTRIENTS: %water, nitrogen, phosphorus, potassium, sulfur, ash LIME: CaCO3 equivalent, input amount POND SOIL: clay, silt, sand, organic matter, pH, phosphorus, calcium, magnesium, potassium, sodium, total nitrogen, ammonium nitrogen, nitrate nitrogen, cation exchange capacity, soluble salts, aluminum, iron, zinc, manganese, copper, sulfate sulphur, lime requirements, free calcium carbonate, exchangeable hydrogen

Table 1. Variables to be measured during the third CRSP experiment cycle

The CRSP data management system was designed to provide a means for storing and retrieving the large number of observations made at the field stations. The current system was developed in late 1985 to replace a small data base system that had been used previously. This paper describes the current data base system and discusses attributes of a good data base system for aquaculture research.

SYSTEM OBJECTIVES AND CONSTRAINTS

The objectives of the CRSP data base system are:

1. To enter data reliably under adverse conditions in locations far from technical support;
2. To have ready access to each station's data so that routine reports and statistical analyses can be prepared;
3. To store all of the data submitted by the stations; and
4. To sort and combine data into formats required for large-scale statistical analyses and modeling.

The primary constraint on developing the data management system was financial; it required the use of existing computer hardware at the various field stations and participating USA universities. Most of the field stations use APPLE IIe® computers with 128K memory, while one or two stations and several of the universities use IBM Personal Computers. The universities also use mainframe computers by Control Data, IBM, Amdahl, and Digital Equipment.[1] Also, any modifications to hardware could not preclude the use of existing software.

DESCRIPTION OF THE SYSTEM

The first task in developing the CRSP data base system was to examine existing aquaculture data base systems. Existing data base systems, such as the Aquaculture MIS (Cyr, no date), are concerned with the everyday monitoring and stock accounting of commercial fish farms. They are inadequate for research projects where large quantities of physical, chemical, and biological data are collected and manipulated. Therefore, the decision was made to develop a new data base management system.

The new system is based on proven commercially available software in order to minimize both costs and difficulties involved in designing new software. Spreadsheets were selected for data entry because they are easy to visualize and use. Also, blanks and errors in magnitude are readily apparent in spreadsheets. MULTIPLAN® was selected because it runs on the APPLE IIe® without requiring system enhancement or expansion. LOTUS 1-2-3® (LOTUS) and MULTIPLAN® were selected for use on the IBM Personal Computers.

The spreadsheet column headings are complete or use unambiguous abbreviations. Column headings also include the units of measurement for the data being entered and an indication of expected magnitude and desired precision (e.g., XX.XX). An advantage of using complete headings is that data reports can be printed directly from the spreadsheets without much additional formatting.

Data can be entered by variable (e.g., all the temperatures for a set of ponds) or by pond (i.e., all of the data for a pond on a given day). Each line in a table contains a complete set of identifiers (e.g., pond number, date, site, experiment number and season). Use of copy commands accelerates entry of these identifiers into the spreadsheets. Additional variables can be included by adding columns to the table. This type of data organization is used by several statistical packages such as the Statistical Package for the Social Sciences (SPSS).

The spreadsheet templates used for data entry closely correspond to the manner of data collection. The eleven spreadsheets are:

1. Daily Weather Measurements
2. Daily Pond Measurements
3. Weekly and Twice Weekly Measurements
4. Diurnal Measurements
5. Fish Stocking, Sampling and Harvesting
6. Plankton and Benthos
7. Occasional Water Quality Measurements
8. Pond Soil Characteristics
9. Pond Morphometrics
10. Miscellaneous Observations
11. Analysis of Nutrients and Lime

Each spreadsheet contains data on all of the ponds at a given station. This arrangement allows each set of observations to be entered without having to switch between several different spreadsheets, each containing the data for one pond.

The field stations are required to submit the completed spreadsheets to the CRSP Management Office. Although the Management Office does not require data summaries and statistical analyses from the field stations, most of the field staff prepare reports for their own use. The sorting and statistical functions of MULTIPLAN® and LOTUS are used in simple statistical analyses.

Diskettes containing the data files are sent from each field station to the CRSP Management Office for review and consolidation. An IBM PC-XT® (PC-XT), equipped with a 20+20 Bernoulli Box (a data storage drive that uses two large, removable diskettes, each with a capacity of 20 megabytes) and with the APPLE TURNOVER® accessory board and accompanying software, is used for the review and consolidation process because it is considerably faster than the APPLE IIe®. The MULTIPLAN® files are read directly from the APPLE-formatted diskettes using the PC-XT floppy disk drive and APPLE TURNOVER® and the files are stored on the PC-XT hard disk. The ® files are then translated into LOTUS for review using the MULTIPLAN® translation function. Files from field stations using IBM or IBMcompatible machines are reviewed directly without the use of APPLE TURNOVER®.

The spreadsheets are well suited to data entry but are not appropriate for filing and retrieving the large amounts of data generated by the CRSP. A high-level relational data base program, RBASE SYSTEM V® (RBASE), was used to create the master data base which contains all of the data submitted by the field stations. To load data into the master data base from the reviewed LOTUS files, six- to eight-digit code names are added to the top of each column of data in the LOTUS files and the multi-line titles are deleted. These modified files can then be read directly into the master data base using he FileGateway procedure in RBASE. Data are moved from the PC-XT to a Control Data Corporation Cyber main frame using a local area network. Data stored on the PC-XT are sorted and reorganized using RBASE, while data manipulation on the Cyber mainframe is accomplished using, a mainframe data base system which is compatible with RBASE. Data tapes in RIM® the formats required by the various CRSP university mainframe computers can be produced on Cyber and sent to the universities for further analyses.

CONCLUSION

The CRSP data base system uses spreadsheet programs to input data from seven field stations located in six countries and then consolidates the spreadsheet data files at the CRSP Management Office using a commercial data base program, RBASE SYSTEM V®. Microcomputers are used for data entry, preliminary analyses and data manipulation, whereas mainframe computers are used for operations involving the entire data set.

The CRSP data base will provide researchers with a tool previously unavailable, a standardized quantitative data set which will become more complete with time. Although the number of field stations is limited, the geographic areas and climates included in the data base are representative of global conditions. It is hoped that other aquaculture researchers will see the benefit of contributing to and sharing the information in this data base. In order to do this, some standardization of data collection and storage is required. This is not to say that all of the contributing researchers should use the same analytical methods, computer systems, or software.

What is required for integrating data sets from many different researchers is the following:
1. Standardized units of measurement;
2. Complete lines of data which include precise location and date of data collection; and
3. Detailed weather records, preferably on a daily basis.

Data sets do not have to be as complete as those collected by the CRSP projects to be of value in the data base. Any researcher wishing further information should contact the CRSP Management Office. By working together, we can better understand the complex processes governing productivity of fish ponds.

REFERENCE

Cyr, B. P. no date. Aquaculture MIS: management information systems for aquaculture. State of Hawaii Aquaculture Development Service, Sea Grant, University of Hawaii, Honolulu, Hawaii, USA.

ACKNOWLEDGEMENTS

The Pond Dynamics/Aquaculture CRSP is funded by grant number DAN-4023-G-55-2074-00 from the Agency for International Development (USAID). The CRSP universities are the University of Arkansas at Pine Bluff, Auburn University, University of California at Davis, University of Hawaii, University of Michigan, Michigan State University, and Oregon State University.

This is the first paper published in CRSP Research Reports. It was submitted in November 1986, and approved for publication In September 1987.

CRSP Research Reports *are published as occasional papers and are available free of charge from the Program Management Office, Pond Dynamics/ Aquaculture CRSP, Office of International Research and Development Snell Hall Oregon State University, Corvallis, Oregon 97331 USA.*

CRSP Research Reports *presents technical papers of research supported by the Pond Dynamics/Aquaculture CRSP.*

Papers are assigned Publication Numbers, which should be included in any request for reprints.

A Guide to Authors may be obtained from the Program Management Office.

[1] The mention of product names does not constitute endorsement or criticism of these products by the CRSP or any of its members.