*Advances in Water Resources Technology, G.Tsakiris (ed.)* © *1991 Balkema, Rotterdam.*
*ISBN 90 6191 184 2*

# HYDRA-PC: A data base system for regional hydrological data management

D.Koutsoyiannis, K.Tsolakidis & N.Mamasis
*Department of Civil Engineering, National Technical University of Athens, Division of Water Resources, Hydraulic and Maritime Engineering, Greece*

**ABSTRACT.** Recent improvements of personal computer capabilities have facilitated the development of computer programs for hydrological data management and processing in order to take maximum advantage of the available hydrological information.

In this paper a software package (HYDRA-PC, HYdrological Data Retrieval and Analysis for Personal Computers) developed for the processing and analysis of daily and hourly hydrometeorological data is presented. The package is made of a number of executable programs and database files. For every hydrometeorological station, the database includes daily and hourly measurements as well as information on the station's peculiarities, quality and accuracy of measurements for an unlimited time period. HYDRA-PC's main characteristics are the rapid entry, updating, retrieval and primary processing of data as well as efficient computer memory and disk usage with the application of special computer programming techniques. The program operates in Greek language and is designed to accommodate the peculiarities of the data collection network (gross inaccuracies in data collection, river stage - discharge instabilities) often encountered in Greece.

## 1. INTRODUCTION

The need of taking maximum advantage of the available hydrological and hydrometric data for a particular region leads to the formation of an organising structure for their storage and analysis. Such efforts prove to be quite useful since primary data provide the most valuable information in hydrology. Good utilisation of such data require the development of sound data banks which can undertake a level of minimum processing, thus providing accurate, and well presented information. The large amounts of data accumulated for hydrological purposes, dictate the use of computer programs for undertaking such tasks.

The extensive use of computers in science in the last two decades has directed hydrologists towards the development of local, regional and national data banks across the world, to facilitate their particular needs, encompassing the particular problems of each region. Such efforts include at national level the *Water Archive System* (UK), the *NAQUADAT* in Canada, the *National Water Data Storage and Retrieval System (WATSTORE)* in the United States, Germany's *Water Resource Data Bank* (Rodda et al, 1985) and at regional and local levels, the *Columbia River Hydromet Management System (CROHMS)* in the U.S.A. (Bissel et al, 1984), the *HYMOS* in The Netherlands (Ogink, 1976, 1981), the *Data Base Management Software for Hydrological Data on Micro-computer* in Belgium (RMIB, 1986), the *FORDATA* in Australia

KEY WORDS: Hydrological data base management, hydrological information

(Goodspeed 1979). In addition, efforts on the exchange and networking of databases have been made (e.g. the *Standard Hydrologic Exchange Format - SHEF*, (Bissel et al, 1984) and guidelines & criteria have been developed for assisting hydrologists in their programming efforts (WMO 1985, Rossi 1979).

The World Meteorological Organization (WMO, 1977) reports that in Greece computer processing and archiving of hydrological data exists through co-ordinated activities between various governmental agencies. However, a centralized national or regional database does not exist to date. The Ministry of the Environment, Planning, Housing and Public Works, the Ministry of Agriculture, the Public Power Corporation and the National Meteorological Service have undertaken the whole effort of collecting and archiving hydrological data, with the last two having progressed in the extensive application of computers. The other agencies, who are responsible for the majority of the hydrological information, keep registers of observations which often include gross inaccuracies or blanks due to the unreliability of the observation network. Furthermore, no primary processing is been conducted and the extraction of information such as discharge from stage data becomes cumbersome and time consuming.

The presented software package, HYDRA-PC (HYdrological Data Retrieval and Analysis for Personal Computers), was designed in the framework of a regional database management system with the possible capability to get linked to a future centralized national data bank. The purpose of its development was to encode in an easily accessible computer database form all historical data required for processing and analysis in the framework of particular research projects (NTUA-MEPPW, 1987, 1988) facilitating their future transmission to a centralized national database.

## 2. DATABASE DESIGN PRINCIPLES

The World Meteorological Organization has proposed a set of guidelines (WMO, 1977) for the development of computerized hydrological databases, which have been adopted at various degrees by the database designers. The above guidelines provide a set of principles and suggestions for the proper data preparation, entry, validation, quality control, update, modification, processing and output as well as the efficient use of computer memory. In the development of HYDRA-PC, WMO's proposals were taken into account.

Like most computerized database efforts, HYDRA-PC was designed to address certain needs and particular problems which occur in Greece. Such needs include the utilization of hourly, daily, monthly and annual hydrologic data (accompanied by information on their integrity) and accurate discharge computations from stage data. The above and other routine database tasks had to be performed in a user friendly environment with ease of operations, speed of processing and expansion capabilities. In addition, data had to be presented in various forms (screen, printer, ASCII file), with the provision of communication with other packages if needed (e.g. HYMOS). The presentation had to include areal values, maxima, minima, average values and the computation of discharge from stage, due to the unavailability of discharge information from the data collection agencies. The package operates in greek language with the need of minimal modifications for the development of an english version.

The principle of *Selective Data Entry* was applied in the development of HYDRA-PC. Data were entered to accommodate accurately certain tasks. For example, for a particular flood computation, hourly raingage and discharge data from a number of days were adequate. The above task was accomplished without the use of the whole annual time series for the particular setting. This fundamental design principle was applied in the construction of records of daily observations as the fundamental data organization unit, as compared to the time series principle, being followed by other programs.

The computer package was developed bearing in mind personal computer limitations. Special provision was made for minimizing storage requirements by packing data in binary form than

using ASCII files. The retrieval of data from accidental loss was accomplished with the separation of pure data files from indexing information and with the possibility of occurrence of runtime errors diminished. In addition, a high level programming language was used (Pascal) to enable complicated mathematical operations, structured data organization, transferability to similar level of computer codes, flexibility and expansion of the program to address future needs. HYDRA-PC's general organizational scheme is presented in Figure 1.

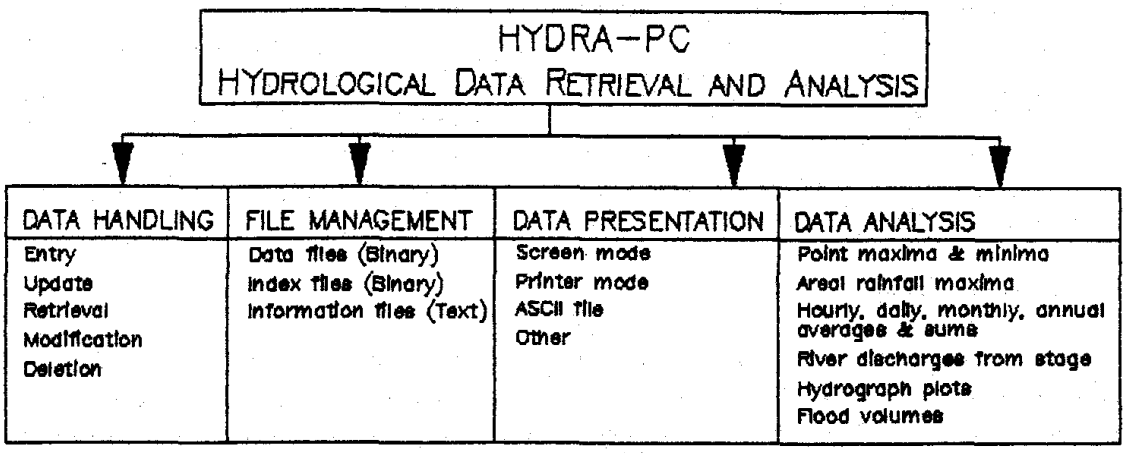


Figure 1. Organizational scheme of the HYDRA-PC package.

## 3. PACKAGE PROFILE

### 3.1 DATA HANDLING

Data handling (entry, retrieval, modification and deletion) is performed through user friendly, menu driven options. Data entry is designed to ease error correction and to speed up keyboard operations, e.g. providing automatic linear interpolation for encoding tape recorded data. Data (raw and processed) are presented in a variety of modes, such as screen, ASCII file and printer. The user can also select from a number of data presentation forms in order to take maximum advantage of the available information and facilitate the linking with other packages.

### 3.2 PRIMARY AND SECONDARY ANALYSIS

HYDRA-PC performs a number of primary and secondary data analysis tasks. Separate computations are made for rainfall and streamflow records. For the case of rainfall data, point maxima are presented and areal rainfall maxima are calculated using the Thiessen method. Monthly and annual average values or sums are also calculated.

Special attention was paid in the computation of the primary hydrological information e.g. river discharge. Daily and hourly river discharges are determined from the corresponding water levels with the use of stage - discharge curves. The method of handling stage - discharge curves is presented further in this paper. For the calculated discharge, average monthly and annual values, as well as maxima and minima for every duration are presented. 72 hour hydrographs can also be constructed from hourly discharge values with the use of elementary graphics. Furthermore, the user has the option of calculating flood discharge volumes having previously defined the desired period.

### 3.3 STAGE - DISCHARGE CURVES

Low volume greek rivers display seasonal stage - discharge instabilities. Because of the above, the experience gained from data manipulation suggests that separate stage - discharge curves not having analytical form should be used for different time periods, with which higher accuracy is obtained, as compared to the use of a single analytical curve made of widely scattered and uncorrelated data points.

For every measurement of stage entered, the program returns the corresponding value of discharge with the use of a modified spline interpolation method. This technique, which is faster than the conventional spline algorithm gives accurate results. It uses polynomials of the logarithmic transformations of each set of points defining a curve. In particular, the polynomials are of the third order between the second and (n-1)th points and of the second order for other ranges. Continuity of the spline curve and of its derivative are being preserved. For the calculation of the curve coefficients, this method uses three or four points which significantly reduces computation time (NTUA-MEPPW, 1987, 1988).

### 3.4 DATABASE FILES

The package consists of a number of executable programs and database files. The executable files handle all program operations. Separate programs exist for the processing of rainfall and hydrometric databases. In addition, for each type of data a corresponding test program checks the structure of the database to avoid organisational errors. Each database consists of a text and four binary files. The text file contains the necessary information for the execution of the program (i.e. stations and basin names, stage - discharge curves, volume of data entered, etc.). Two binary data files contain the corresponding hourly and daily data and two binary index files include the necessary information for ordering and retrieving the hourly and daily data, since the data files are serially expanded according to the entry sequence.

### 3.5 DATA FILE ORGANIZATION

HYDRA-PC's binary data files are organised in records. Each record contains information for one station and for one day and is made of a number of fields. A daily record consists of six fields which store information on the status (valid or deleted entry), date, station code, magnitude and quality of measurement and on the existence of hourly data. An hourly record consists of four fields which store information on the status, date, station number and measurements (an array of 24 values).

### 3.6 INDEX FILE ORGANIZATION

The records of the database under consideration were indexed following to the *B+ tree* principle. The B+ tree structure (Bayer et al, 1972), which utilises the logarithmic search technique is a commonly used and efficient indexing method. The structure and ordering rule of B+ trees is illustrated in Figure 2. As compared to *Balanced tree* indexing, B+ trees are characterized by high performance in sequential[1] and direct accesses[2], in growth rate[3], but low in volatility[4] (Hanson, 1988).

B+ tree structures are made of several *levels*: the *root page*, the *internal* and the *leaf* pages. The number of levels of a B+ tree is called the *height* of a tree. Each page contains several *items*. The item, which is the fundamental unit of a B+ tree, is made of a *key string*, a *data reference*, and a *page reference*. The key string is used for accessing purposes. The data reference points to the location where the data record associated with the key is found and the page reference forms the link in the B+ tree by pointing to a page where all keys are greater than the current item's key (Borland, 1987). In HYDRA-PC key strings include the date of data entry

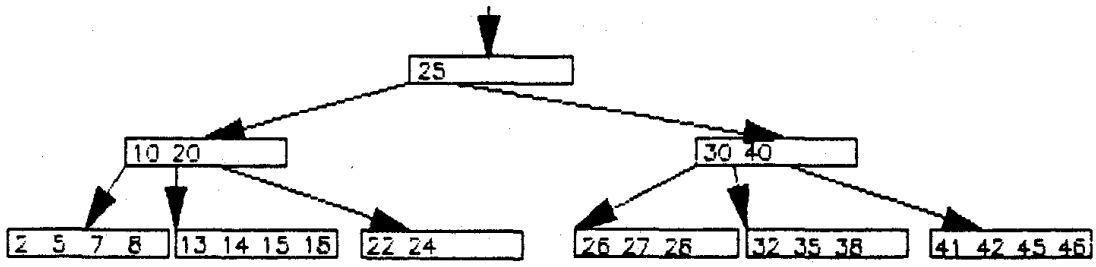


Figure 2. An illustration of the B+ tree ordering rule (from Wirth, 1976). The order of the B+tree is 2. For every parent page with k items, exist (k+1) child pages. For example, the internal page with items (10,20) has 3 children. Its first leaf page (2,5,7,8) contains items of magnitude less than 10; the second (13,14,15,16) contains items between 10 and 20, etc. The search path to find e.g. 35 in the structure is: [root] -> (greater than 25) -> [right child page of level 1] -> (between 30 and 40) -> [middle leaf page of level 2].

and the station code. The hierarchy of ordering of keys adheres to the following: a) between two keys, greater is the one with the most current date and b) between two keys of the same date, greater is the one with the greater station code. Both, key strings and page references are generated internally by the program.

In a B+ tree of *order n*, a) all pages contain between *n* and *2n* items, thus providing store utilization at least 50% - since pages are always at least half full - (Wirth, 1976) and b) an item search can be successfully executed with less or equal to $\log_n N$ page searches, where N is the existing number of items (Kollias, 1986). Most applications of HYDRA-PC use B+ trees of order 24 and height 3 or 4, thus requiring a maximum of about three searches per item. Items are distributed evenly in the structure, i.e. they are balanced. Adding and deleting items causes a temporary unbalancing, which is addressed by appropriate routines. It should be also noted that large numbers of items can be placed in a B+ tree in a small number of levels; therefore the search path length remains short and uniform for most items.

### 3.7 MEMORY USAGE

Although the cost of microcomputer components is decreasing significantly, large amounts of costly space are still required to run hydrological databases. HYDRA-PC is a program that utilises efficiently both, computer memory and data storage (disk) space.

As mentioned above, data and index files are of binary type. Binary files occupy substantially smaller amounts of memory space and perform read / write operations faster, as compared to ASCII or text files.

For the purpose of further reducing memory requirements, HYDRA-PC "packs" data prior to their storage. This is achieved with the conversion from floating point to ordinal type

---

1. *Average Sequential Access Time* is a measure of the time required to access a record when a file is being processed sequentially (Hanson, 1988).

2. *Average Direct Access Time* is the average time it takes to access a record (Hanson, 1988).

3. *Growth Rate.* In allowing space for a a file, it is essential to know whether the file is growing in size and if so by how much. In planning for the expansion of a file the strategy adopted for deletions is important (Hanson, 1988).

4. *Volatility* is a measure of of the changes in a file with time. It is expressed by the ratio (Hanson, 1988):
(number of additions + number of deletions) / (number of records in the file at the start)

(integer), thus resulting to a significant gain (about 50%) in disk memory. For every rainfall data record a total of eight bytes is used; for every daily hydrometric measurement nine bytes are used and if hourly data exist, a total of fifty five bytes per day are needed.

It should be also noted that in order to manage memory efficiently, the *dynamic memory allocation* technique was utilised so that only the necessary information for running the program is loaded at any time. In particular, this proved to be quite beneficial with the stage - discharge curves where only the curves of the station under consideration are loaded at any time (NTUA-MEPPW, 1987, 1988).

## 4. CONCLUDING NOTE

HYDRA-PC is a flexible and user friendly database management system with considerable expansion capabilities because of its programming and indexing structures. The package makes efficient use of personal computer memory and disk space. Special provisions were made to accommodate peculiarities of the greek setting, thus provide accurate data. It is currently operated by government agencies to manage hydrological data of large regions in continental Greece.


### ACKNOWLEDGEMENTS

The development of HYDRA-PC was funded by the Greek Ministry of the Environment, Planning and Public Works, in the framework of the research projects *Hydrological Investigation of the Thessalia Water Basin (1987) Project No. 8764700* and *Appraisal of Existing Potential for Improving the Water Supply of Greater Athens (1988) Project No. 8576710,* which were undertaken by the National Technical University of Athens, Division of Water Resources, Hydraulic and Maritime Engineering. The authors acknowledge the valuable assistance of the Projects' Director, Professor T. Xanthopoulos.


### REFERENCES


Bayer, R., McCreight, E., (1972), 'Organisation and Maintenance of Large Ordered Indexes', Acta Informatica, 1(3)173-189.

Bissel, V.C., Pasteris, P.A., Bennet, D.G., (1984), 'Standard Hydrologic Exchange Format (SHEF)', Journal of Water Resources Planning and Management ASCE, 110(4)392-401.

Borland International Inc., (1987), Appendix B: B+ Tree Structure, Turbo Database Toolbox Owner's Handbook, California, U.S.A.

Goodspeed, M.J., (1979), 'A Database for Hydrologic Time Series Data', Proc. Hydrology and Water Resources Symposium, Institute of Engineers, 99-100, Perth, Australia.

Hanson, O., (1988), Design of Computer Data Files, 2nd ed., Pitman Publishing, London, U.K.

Kollias, J., (1986), Data Structures, Simmetria Publishing, Athens, Greece, (in Greek).

National Technical University of Athens , Division of Water Resources, Hydraulic and Maritime Engineering and Ministry of the Environment, Planning and Public Works (NTUA-MEPPW), (1987), 'Computer Programs for Hydrological Data Archiving and Processing', Hydrological Investigation of the Thessalia Water Basin, Project No. 8764700, Vol.1, Athens, Greece, (in Greek).

National Technical University of Athens , Division of Water Resources, Hydraulic and Maritime Engineering and Ministry of the Environment, Planning and Public Works (NTUA-MEPPW), (1988), 'Computer Programmes for Hydrological Data Archiving and Processing',Appraisal of Existing Potential for Improving the Water Supply of Greater Athens, Project No. 8576710, Vol.5, Athens, Greece, (in Greek).

Ogink, H.J.M., (1976), 'Development of an Integrated Hydrological Model System', Proc. Exchange Meeting Vituki-Delft, Delft Hydr. Lab. and Budapest Inst. Water Resources (Vituki).

Ogink, H.J.M., (1981), 'Hymos - A Data Processing System for Hydrological Time Series', Proc. International Conference

on Numerical Modelling of River, Channel and Overland Flow for Water Resources and Environmental Applications, Bratislava, Yugoslavia.

Rodda, J.C., Flanders, A.F., (1985), 'The Organization of Hydrological Services', Ch.14, II, Facets of Hydrology, John Wiley & Sons, New York, U.S.A.

Rossi, G., (1979), 'Hydrometeorological Data-Survey and Automatic Processing', Proc. Design and Management of Systems for Field Data Collection and Processing (Subject A.d.), International Association for Hydraulic Research, 487-495, Cagliari, Italy.

Royal Meteorological Institute of Belgium (RMIB), (1986), Data Base Management Software for Hydrological Data on Micro-computer - Explanatory Notice -, G06.3.01 (micro), Brussels, Belgium.

Wirth, N., (1976), Algorithms + Data Structures = Programs, Prentice-Hall Inc., Englewood Cliffs, New Jersey, U.S.A.

World Meteorological Organisation, (1977), Statistical Information on Activities in Operational Hydrology, No.464, Geneva, Switzerland.

World Meteorological Organisation, (1985), Guidelines for Computerized Data Processing in Operational Hydrology and Land and Water Management, No.634, Geneva, Switzerland.