TN-100

HYDROLOGICAL DATA PROCESSING

TECHNICAL REPORT

ON

TRAINING

OF

HEMANT CHOWDHARY

UNDER

WAMATRA - II

NATIONAL INSTITUTE OF HYDROLOGY

ROORKEE

FEBRUARY 1992

PREFACE

In December 1985 the governments of The Netherlands and India agreed on an Indo-Dutch Training Programme on Water Management (WAMATRA). Phase I was concluded in 1986-87. Phase II of WAMATRA started in August 1990. National Institute of Hydrology was associated along with other institutions in this project in a few areas. One of the items was "Data Storage and Retrieval System". The Delft Hydraulics from the The Netherlands side was to organize courses, training and workshop under this item.

In April/May 1991 a one week course was organized by Prof. H J M Ogink and Mr. Johan Crebas of Delft Hydraulics. A software package on Storage, Processing and Retrieval of Hydro-meteorological Data called HYMOS was introduced to the participants. After this introductory course Shri Hemant Chowdhary recieved an in-depth training on the software at Delft Hydraulics, De Voorst for two months during Oct./Dec. 1991.

This technical report on his training on HYMOS covers general principals of database management system in its Part I and a detailed information of the capabilities and procedures of HYMOS in Part II. An idea has been given, in Part I, about the necessary features of a good database management system. While in Part II various capabilities of storage, processing and retrieval of hydro-meteorological data have been elaborated. An attempt has been made to discuss every option of software in Part II so as to give a complete understanding of the package. The main components of the package are entry and editing (storage of data), validation, completion, transformation, statistical analysis time series analysis (processing of hydrological data) and reporting, retreival and transfer of data (retreival of data).

Roorkee, February 92.

Satish Chandra

Satish Chandra

CONTENTS

		PREFACE	(i)
		PART I GENERAL PRINCIPLES OF DATA PROCESSING	
1.0		INTRODUCTION	I-1
2.0		DATA OPERATIONS	I-2
3.0		COMPUTER STORAGE AND PROCESSING	I-7
	3.1	Hardware	I-7
	3.2	Programming Languages and Computer Software	I-12
4.0		NATURE OF DATA	I-14
	4.1	Data Types	I-15
	4.2	Computer Storage Systems	I-16
	4.3	Coding of Data	I-17
5.0		DATA BASE DESIGN	I-18
	5.1	The Role of DBMS	I-18
6.0		OPTIONAL ASPECTS OF DATA PROCESSING	I-19
	6.1	Accuracy of Keyed Data	I-19
	6.2	Accuracy of Chart Data	I-20
	6.3	Validation and Updating of Data	I-20
7.0		DATA RETRIEVAL	I-20
		Figures of Part I	I-F-1 to I-F-4
		Tables of Part I	I-T-1 to I-T-2
		PART II A SYSTEM FOR STORAGE, PROCESSING AND RETRIEVAL OF HYDRO-METEOROLOGICAL DATA (HYMOS)	
1.0		GENERAL OVERVIEW	II-1
	1.1	Computer Configuration Requirements	II-1

	1.2	HYMOS in a nutshell	II-2
2.0		TYPES OF DATA	II-7
	2.1	Overview	II-7
	2.2	Catchment data	II-7
	2.3	Station data	II-8
	2.4	Geo-hydrological profiles	II-11
	2.5	Equidistant time series	II-11
	2.6	Non-equidistant time series	II-19
	2.8	Combined data and parameters	II-21
3.0		Structuring the database	II-24
	3.1	Database directories	II-24
	3.2	Extent of a database	II-24
4.0		Data preparation for transfer	II-25
	4.1	Introduction	II-25
	4.2	Equidistant time series	II-25
	4.3	Non-equidistant time series	II-37
	4.4	Stage-discharge data	II-39
5.0		Data Validation	II-40
	5.1	Screening	II-40
	5.2	Time series graphs	II-42
	5.3	Relation curves	II-46
	5.4	Double mass analysis	II-47
	5.5	Series homogeneity tests	II-49
	5.6	Spatial homogeneity test	II-50
6.0		Series completion and regression	II-53
	6.1	Interpolation	II-53
	6.2	Regression analysis	II-55
	6.3	Rainfall-runoff simulation	II-56
7.1		Processing of flow measurements	II-58
	7.2	Fitting of Rating Curve	II-58
	7.3	Validation of Rating Curve	II-63
	7.4	Extrapolation of Rating Curve	II-64
	7.5	Stage discharge Transformation	II-65
8.0		DATA COMPILATION	II-65
	8.1	(Dis-) Aggregation of Series	II-65
	8.2	Series Transformation	II-66
	8.3	Minimum and Maximum Search	II-68
	8.4	Areal Rainfall	II-68

	8.5	Kriging	II-69
	8.6	Computation of Evapotranspiration	II-69
9.0		STATISTICAL ANALYSIS	II-70
	9.1	Basic Statistics	II-70
	9.2	Fitting Distributions	II-71
	9.3	Statistical Tables	II-71
	9.4	Generation of random variables	II-72
	9.5	Frequency and duration curves	II-72
10.0		Time series analysis	II-73
	10.1	Introduction	II-73
	10.2	Correlation analysis	II-73
	10.3	Spectral analysis	II-75
	10.4	Range analysis	II-76
	10.5	Run analysis	II-77
	10.6	Storage Analysis	II-78
11.0		REPORTING AND RETRIEVAL	II-79
	11.1	Reports	II-79
	11.2	Mixed tables	II-82
	11.3	Graphics	II-83
	11.4	Retrieval and transfer	II-84
		Reference	II-86
		Figures of Part II	II-F-1 to II-F-22
		Tables of Part II	II-T-1 to II-T-12

PART I

GENERAL PRINCIPLES
OF
DATA PROCESSING

PART I

GENERAL PRINCIPLES OF DATA PROCESSING

1.0 INTRODUCTION

The increasing capability and availability of computer related technology is having a revolutionary effect upon the techniques available to those responsible for the assessment and management of natural resources. For many years there has been a continual and rapid growth in database systems, not only with respect to the sheer number of such systems, but also with regard to the amount of information being stored and the complexity of applications being developed. The most fundamental advantages of computerized processing systems and perhaps still the most significant are:

- Data are stored in a compact, organized manner;
- Data may be rapidly retrieved in a wide variety of formats and aggregations.

With respect to subsequent data applications there are other important benefits. The computer's ability to perform systematic consistency checks offers a powerful aid to improving and maintaining the quality of data and data analysis techniques may be much more sophisticated than those found in any routine manual processing system.

The many potential benefits of computerised data processing may only be fully realized by proper planning and management. This involves equipment (Hardware), computer programs (software) and personnel. Unfortunately, there are still problems of compatibility between different computer, and this incompatibility affects both hardware and software. Care needs to be taken at all stages of system development in order that such problems may be minimized.

A major constraint in developing countries is the shortage of highly trained professionalsand technicians. By putting mechanical and logical power, a computer, into the hands of those people who are available, their effectiveness may be greatly increased. The obstacles to this process are usually the inexperience of local computer suppliers of finding suitable personnel to be trained for providing "in-house" support. However, where these problems have been overcome, computers have been found to facilitate greatly the transfer of technology.

The intent of the database organization is to provide a large, integrated store of data that any user can tap as needed. Data should be defined and stored independent of the limitation of any specific program. Related data in various areas of the files must be able to be created, retrieved, updated, or deleted. Elements of the data base are the data collection itself, direct access storage devices,, the data element dictionary, data base administrator and data base management.

2.0 DATA OPERATIONS

Various data operations may be categorized into three groups. The first group of operations is associated with data collection, and covers the stages between data observation and its arrival at some processing centre. The second group of operations relates to data processing and covers the preparation, input and quality control of the data, the updating of the database (or data bank), techniques to enhance the utility of the data, and standard data retrieval options. The final group of operations are those associated with data application, and involve the use of data in subject specific analyses such as the estimation of aquifer recharge, land use evaluation and storage requirement for a multipurpose reservoir. The first and the last groups fall outside the scope of this publication and hence would not be dealt in details.

a) Data collection

The components of data collection are the sensing of a required element or parameter (hereafter called parameter), the recording of the value sensed, and the subsequent transmission of this value, or groups of values, to the data-processing centre. Table 1 shows the relationship of the various operation, and gives examples of each. Where analysis are to be performed in a laboratory, a physical sample must also be transferred from the field.

The two concepts which relate data collection and data processing are:

(i) The principle of redundancy: Data redundancy involves the over-definition of certain parameters by the input data. The redundant data is used by the processing system for validation purposes. System redundancy requires the duplication of sensors, recording systems, complete observation stations, and transmission and processing devices. The extent to which either or both types of redundancies are practised relates to the design objectives of the processing system.

(ii) Initial data validation should remain, as far as possible, the responsibility of those who collect the data. This is particularly valuable in the identification and correction of data errors.

b) Data Processing:

It is the group of operations to be performed in data processing which are the focal point of this publication. The data collection exercise provides a variety of text descriptions, numerical values, charts, computer compatible media, etc. It is the task of the data processing system to convert these diverse data, recorded on diverse media, into integrated data sets structured to suit the information requirements of the data user. Table 2 shows the main components of a data-processing system.

b.1 Data Preparation: This activity comprises the operation necessary to convert data from the format in which it is received to a format suitable for input to the computer. The complexity of these operations is obviously governed by the degree of computer compatibility of the recording media. Much handwritten data may need to be transcribed and perhaps replaced by some coded value. This produces punching documents (sometimes called coding sheets) from which the data are subsequently key punched directly onto the computer or onto some compatible storage media, e.g., punched cards, diskettes or magnetic tape.

This preparation is both tedious and liable to produce transcription or coding errors. Methods for reducing or eliminating data preparation effort may be employed to overcome this difficulty. These methods include the design of field data sheets from which keying may be performed directly.

b.2 Data Entry : There has recently been a major shift in methods of data entry (input). Punched cards and paper tapes have been widely replaced by key to tape or key to diskettes systems. Large scale keying operations normally comprise two stages; in the first (punching) stage data is keyed by one operator, in the second (verifying) stage the same data in keyed by another operator and compared with the original input. Any differences between the two data sets cause the keyboard to lock, and the data value being entered can be checked. This system is very successful in eliminating data entry errors.

There are other more specialized data entry methods, e.g. the conversion of charts and maps to digital format using manual or automatic digitizers.

b.3 Data Validation: Once entered into the computer, data should be subjected to a set of checks designed to identify incorrect or unusual data values. These may be simple. Checks that compare the value entered with an expected range for that parameter, or more complex ones that compare, say, daily rainfall totals at one site with totals, recorded at adjacent sites. If data has been coded, code values and code combination can be checked for validity.

Mis-coded or mis-punched data identified in this way can normally be easily corrected. However, correction of suspect data values, or the querying of missing values need to be referred to those responsible for data collection, an important reason for allowing this initial phase of data processing to be carried out by the field observers themselves.

b.4 Primary Processing : This stage of processing is concerned with preparing and assembling the data in the format necessary for it to be added to the existing data base. This may include standardization of measurements units, additional levels of coding for storage purposes, and the estimation of derived parameters, e.g. estimation of Penman evaporation from climatological data. The amount of primary processing necessary is related to the degree of coding of the data before input. Systems which utilizers less initial coding are easier for the computer user, but need more powerful computer facilities, and vice versa. The extent to which manual coding is used is governed by the quality and availability of data preparation and computer programming staff, and the capabilities of the computer system used.

b.5 Data base Updating : Having processed the data to the correct format, it may be incorporated into the data base using an updating program. It is usual to update the data base at some fixed time interval, monthly being a typical interval for natural resource planning data. This requires the allocation of space on the various storage media to hold recently input data until the next update run.

The data base comprises sets of data cross-referenced in a way which reflects the relationships between various data item, and data groups. This structuring of data sets is fundamental to the proper design of any data base.

Table 2 shows, as is traditional, that data bases updating does not occur until the 5th of eight stages. With this arrangement, validation is done on every "transaction" before it is added to the data base. A recent development associated with widespread interactive processing facilities is for data base updating to be the first stage as part of data entry. Subsequent stages then operate, on the data base using its facilities for manipulation and maintenance of data integrity. With this arrangement both management and users may apply the same validation tests, which can be helpful when user's demands for urgency exceed management resources for validation. This use of full data base facilities in regional, even field offices has become feasible with microcomputer software structured identically to mainframe software which supports the national data base, although having capacity for less data.

b.6 Secondary Processing : This stage of processing covers the reports generated and analyses performed on a routine basis after the data base has been updated. For example, it is usual to provide statistics comparing the most recently input values of a parameter with its long-term variation, in order that any trend in values can be monitored.

There are also methods of aggregating or interpolating data, including the in filling of missing data values, which generally enhance the subsequent usefulness of the data, and program to perform these analyses are normally run at this stage.

b.7 Retrieval

An important objective of an automated data processing system is to provide a rapid and comprehensive response to ad hoc requests for data retrieval and interpretation. Thus, a wide range of retrieval options should be available to allow the interrogation of the data base and the selection of ranges and combination of data, e.g. the selection of all values of parameter x above a value V which were recorded at location Z between dates A and B.

Retrieval options should also include the facility to combine selected data sets and prepare data files for subsequent input to standard statistical software or user application programs. Another attribute of retrieval options is the ability to allocate any suitable computer derive for the presentation of the selected data.

b.3 Output : Modern computers support a wide range of output devices. These include, in addition to the range of computer storage media, printers, plotters, and cathode ray tubes or visual display units (CRTs or VDUs - almost identical to a conventional television screen). Of increasing use are devices supporting both graphics and colour options.

As already stated, it may be desirable to split the processing operations between physical locations. It is proposed that initial data preparation and validation are best performed by field staff as a continuation of the data-collection exercise. The use of microcomputers should be considered to assist in this function. However, the computer and manpower requirements for the operation of an integrated water data base imply the existence of a central computer facility.

Initial data preparation and validation by field staff using microcomputers as a continuation of data gathering has been implemented in some countries. In such cases, a central computer facility may not be necessary when the advantages of integration are achieved through use of a common data structure to allow free interchange of data between computers. It has recently become feasible to implement on microcomputers a relational data base which supports all the functions provided on the largest computers for resources data although for smaller qualities.

c) Data Application :

The objective of many automated data processing systems, particularly at an early stage of development, is simply to create a readily accessible archives of homogeneous data which has been subjected to validation procedures. This, in itself, is a significant objective, given the nature and diversity of environmental data.

However, the purpose of collecting data is to convert it into information which will assist the planning, design and operational functions of an organization. The conversion processes from data to information are the data applications, which may or may not be computerized. Data base applications may broadly be divided into routine and non-routine.

Routine applications comprise the continuous data processing and report generation cycle fundamental to most automated systems. Examples are: the weekly estimation of soil moisture status for irrigation scheduling, the monthly estimation of groundwater recharge from rainfall and river flow, and the preparation of an annual year book of climatological or hydrological statistics.

Non-routine data applications are generally related to the specific limited objective of a project or piece of research, and may themselves be divided into two categories, those using standard commercially available statistical or mathematical software packages (e.g., Statistical Program for the Social Sciences (SPSS), Statistical Analysis System (SAS), Mathematical Programming System (SPSX), and those of a more specialized nature which require user written software (e.g. catchment simulation models using rainfall, evaporation, groundwater level and runoff data).

3.0 COMPUTER STORAGE AND PROCESSING

The computer is able to execute a few extremely simple operation (instructions) very accurately, and very quickly. Thus, by breaking any qualifiable problem down to a sequence of simple arithmetic and logical operations it may be analysed using a computer. It is very essential to have some understanding of the nature of computers since an automated data-processing system may not be conceived without a computer. Broadly, a computer system may be divided into two component : hardware the physical apparatus which constitutes a computer, and software - the sets of instructions (programs) necessary to operate the computer and process data.

3.1 Hardware :

Hardware comprises three groups of devices :

a. Central Processing Unit (CPU)
b. Mass Storage (or backing storage or secondary storage)
c. Input/output devices

The importance of the CPU is reflected in the fact that all other hardware units are considered subordinate, and are given the general name, peripherals. The relationship between these hardware groups, and the principal components of each, are shown in Figure 1 .

3.1.a Central Processor Unit

The central processor is the heart of a computer system. It stores program instructions, stores the necessary data, and brings these two sets of information together to execute a program. The major classification of computers is determined by CPU characteristics. These classes include large 'mainframe' computers, 'mini' computers and 'micro' computers. The essential difference between these machines is speed, core size and processing capabilities of the CPU.

CPU has four main components:

a) Core memory (or main memory or primary storage);
b) The control unit;
c) The arithmetic and Logical unit;
d) The machine instruction sets.

3.1.b Mass Storage Devices

Core is expensive, much more expensive per byte of data stored than other computer storage media. A balance must be reached between expensive,. fast access memory, capable of addressing individual words and chapter, slower access storage which normally can address only larger data blocks.

Thus, the total storage available in a computer system generally has two components; core (or RAM), containing data and program instructions currently being executed, and backing store, to and from which data sets or programs are copied. The backing store has a much larger storage capacity than core, and for this reason the peripheral units involved are called mass storage devices. Because these devices are directly under the control of the CPU they are said to be on line.

There are other important differences in the nature of core and mass storage which can be used to advantage. Core is only a temporary storage, when the computer is switched off, the contents of the storage are lost. Mass storage, however, uses permanent storage media in which data and program can be stored for indefinite periods. Here the data is encoded in the same way as it is represented in core, facilitating high speed transfer into core for processing purposes.

There are two main media used for mass storage, magnetic disks and magnetic tapes. A feature common to both is their prodigious storage capacity, in the order of 20-200 Mb (Mb = Mega byte = 10 ** 6 bytes), compared to say 0.5 - 2.0 Mb core storage for a typical machine.

Another important advantage of these mass storage media is that they are exchangeable, i.e. can be removed from their peripheral units and replaced by others. Thus, an infinite number of data sets and program can be made available to the CPU for processing. The only limit is the number of tapes and disks which can physically be stored adjacent to the computer.

3.1.c Input/output devices:

The previous paragraph considered the hardware necessary to control and execute data processing, and to store data and programs. It is apparent that there must exist various methods of entering and extracting data from the computer. These duties are performed by input and output devices.

(i) Input Devices:

The main groups of input devices have already been listed in the table 2 under the heading Data Entry. In addition to data already recorded on computer compatible media, there are four groups; key punching from source documents, automatic reading of data forms, digitizing of maps and charts and electronic communication links.

(a) Key Punches: A key punch is a generic name for a device through which data is entered on a keyboard, very similar to a typewriter, translated into a computer readable code and then stored on some computer compatible media. Most modern key punches are fitted with a VDU to allow visual checking and editing of data as it is keyed. The computer compatible media produced depends upon the type of key punch installation and may be punched cards, punched paper tape, or some form of magnetic media (e.g. tapes and disks). In some key to disk input systems, the disk is actually part of the computer on-line mass storage. Indeed, on small machines, and where small volumes of data are concerned, this method of direct data input through a VDU is the most common and convenient. However, in large installations, particularly where significant volumes of data are being handled, these key punching operations are performed off-line, and the data stored on computer compatible media as an intermediate stage before being entered into the computer.

(b) Automatic readers: These are devices which read specially marked or printed forms. In the first category are optical mark readers (OMR) or mark sense readers which can read pre-printed cards taken to the field and marked, normally with a soft pencil, at predetermined positions on the card, to signify the value or condition of the parameter. Other systems can actually read characters directly. They are optical character recognition (OCR) and magnetic ink recognition (MICR) systems. Some OCR systems are able to read neatly handwritten numeric values directly, but in general both require the printing of characteristics using specific type fonts. A final category of automatic devices is voice readers, able to interpret values dictated onto a cassette.

The method of most interest for water data purposes are OMR, and OCR reading of handwritten values. However, both methods are currently relatively expensive and unreliable, and it is considered that their use in developing countries would be difficult at present.

(c) Digitizers: If the data to be entered are in the form of charts or maps, then the use of a digitizer should be considered. This machine consists of a board on which the chart or map is laid, and a cursor which is used to follow the lines on the chart or map. The x-y co-ordinates of the lines are recorded at some interval which is selected either manually, depressing a button to store the current x-y value of the cursor, or automatically. Chart reader systems are required to record one or more simple pen tracings across a chart.

A final class of digitization techniques relate to the conversion of photographic or other remotely sensed data which is continuous in space. In these cases specially designed video cameras and "Zoom transferoscopes" may be used to produce digital images.

(d) Telemetry: Data could directly be transmitted, by electronic/telecommunication techniques called telemetry, to the central station without prior recording in some form at the place of data collection site itself. Practically all telemetry systems have the same components, an electromagnetic signal representing the value of some required parameter, a carrier signal and a transmitting medium through which the combined ("modulated") carrier and data signal is sent. The data signal is typically a square waveform bit pattern. The carrier may be a UHF/VHF sine wave. the transmittal medium is cable, for telephone based telemetry links, or air for radio links.

ii) Output devices:

For output which needs only visual inspection, the VDU is widely used. However, most output is needed in a printed form (Hard copy), for which purpose a large variety of printers exist, and printed graphical output is obtained from a equally wide range of plotting devices. If the data is required at some other location to which a telecommunication link exists, telemetry may be used. A final category of output devices widely used in the sphere of documentation is the automatic production of computer output on microfilm (COM)

(a) Visual Display Unit: Visual display units (VDUs or Cathode ray tube CRT) are identical in appearance to a conventional television screen and can contain between 200-5000 characters of data. Alternatively, graphs and maps with shading and with colour may be used. An example of this later use is the display of rain radar information where rainfall intensities are colour coded onto a locational base map of the area surrounding the radar installation.

This type of terminal is extremely useful for data validation and retrieval when only limited but rapid access to the data is required, i.e. to check the contents of an input data file before program execution, or to check a result file before printing. A VDU is almost invariably used as the control terminal of a computer, and is widely used for data entry.

(b) Printers: The main division in printing devices is between line printers, which as the name suggests print a line at a time, and character or serial printers in which characters are printed one by one. Line printers are devices used for printing large quantities of output. One or more lines are printed at a time, and speeds vary from 300 to 3000 lines per minute. Basic graphical output called line printer plots may also be produced by a line printer but resolution is low and since the direction of paper feed cannot be reversed everything must be plotted at one time, making programming more difficult.

A number of small printers are now available which produce reasonably good quality graphics using the dot matrix principle. Early example of these used an electrostatic system but impact dot matrix heads are now also used. Ink jet printers offer the possibility of coloured graphics. As these printers can produce both diagram and text and are now relatively inexpensive, system cost is considerably reduced.

(c) Plotters: Unlike printers which use a fixed set of characters printing in fixed positions, a plotter (or x-y plotter) is a device that moves pen in a set of x-y co-ordinates. The machine is driven by the computer and enables maps, graphs and diagrams to be prepared.

There are two main plotting systems, identified by the way in which the plotting surface is mounted. The first is the flat bed plotter where x-y axes are in one plane, the pen moves in two directions and the plotting surface is stationary. The second is the drum plotter where the x axis pen movement and the y axis movement is generated by the movement of the paper beneath it. The paper is in the form of continuous stationary mounted over a drum which can rotate in either direction.

(d) Telemetry: If the data are to be output through a telemetry system, some communications hardware/software is necessary to encode and transmit it. The main problems, as with all telemetry systems, are ensuring a compatibility of data coding, and maintaining the reliability of transmission.

(e) Microfilm: Computer output to microfilm (COM) is a method suited to the needs of systems generating very large amounts of documentation. This is seldom true of natural resources data systems. It is attraction to other commercial world because of its high speed, upto ten times faster than a line printer. However, its expense and sophistication make it generally unsuitable in developing countries.

3.2 Programming Languages and Computer Software

Computers can only operate by responding to a set of instructions. Such a set of instructions is called a program, and these are written in one of several programming languages which the computer can translate. To contrast the nature of programs with that of the computer hardware, they are given the generic name of software. It is the combination of hardware and software which provide the computer resources.

3.2.1 Programming Languages

Machine language is the most basic method of communicating instructions to the computer. However, this language is difficult for the computer user to understand, so initially numeric codes were replaced by mnemonic codes that were easier to write and remember. This level of languages still in the same format as machines language, is called assembler, and all computers come with their own assembler languages.

The problem with the assembler language is that it is very machine dependent and requires each machine operation to be specified. From the roof of this low level language have grown a large number of higher level languages whose main objective have been:

(a) To make program writing easier, i.e. as near as possible a mixture of English Language instructions and standard mathematical notation,

(b) To make programs as transferable is possible, i.e., the language used is independent of the type of computer on which the program is run.

The development of database systems operating on large quantities of highly structured, interrelated data requires integrated programming environments. Database programming languages combine the facilities for database definition and manipulation of data model with the data structures and operators of a programming language. Most current database programming languages are an integration of Pascal or Modular like languages and versions of the relational database model.

Commercial database management systems typically provide interfaces to the common programming languages such as COBOL, FORTRAN, Pascal, and PL/1. These interfaces may take the form of a library of database access procedures which may be called from programs written in high, level languages. Such programmes are then compiled and linked to the appropriate precompiled database procedures.

In recent years a number of database programming languages, including Pascal/R (Schmidt, 1977, Schmidt and Mall, 1980), Modula/R (Reimer, 1984), PLAIN (Wasserman et.al., 1979, 1980), ASTRAL (Amble et.al., 1976), RIGEL (Rowe and Shoens, 1979), Theseus (Shopiro, 1979), and RAPP (Hughes and Connolly, 1987), have been developed which integrate a complete set of relational database operations into a well structured, strongly typed language. Although the design goals vary among these languages, they are all united in their main objective, which is to provide an effective tool for the development of database application program.

3.2.2 Software:

Software is the generic name for programs which are written to direct the operation of the computer and make it perform specific data processing and analysis tasks.

There are many categories of software but the most relevant groups are:

(i) Operating systems,(ii) Compilers, (iii) Communication Systems, (iv) Utilities, (v) File processor, (vi) General application packages, (vii) Specific application software.

Item (i) and (ii) are essential feature of the computer and are supplied by the computer manufacturer. Item (iii) is also essential if the computer is required to have any communication links to other sites. Items (iv) and (v) are options supplied by the computer manufacturer or by an independent software company and their purpose is to improve the management and manipulations of data sets. These items usually need to be written specifically to match the characteristics of the hardware on which they will be run (machine dependent). Item (vi) and (vii) are designed to be used on as wide a range of hardware as possible and are oriented to assisting in data analysis.

Softwares for mainframes and minicomputers is normally brought or leased direct from the computer manufactures. This software is expensive because of the high costs of program development, but is generally of good quality in terms of reliability, documentation and after sales support. Conversely, microcomputer software is extremely cheap but this may be reflected in the quality of the software itself or of the documentation.

4.0 NATURE OF DATA

There are certain universal characteristics of data, which are independent of the quality or element measured, but strongly influence data collection and storage. Two important characteristics of data are the distribution of the measurements in space and time. In water systems there are variations in the values of physical parameters in both space and time.

When monitoring natural resources data, measurements of quantities of interest have to be made at a number of sampling points and when time variation is also involved a series of observations at different times are also needed. Usually to economize on data-collection costs the observations have to be spaced as widely as possible consistent with being able to interpolate intermediate values to an acceptable accuracy. Three dimensions may be recognized : the place of the observation, the time it was made, and actual parameter(s) observed. Figure 2(5.1) shows this three-dimensional nature of data. The entries in the matrix may be the numerical value of some measurement, a quality or property of the place, or some text note or comment.

Thus with every observation of a parameter are associated other information such as the time and place of measurement and the name of the parameter observed which serve to identify the data. A data storage system has to reflect this identification information in some manner so that the data can provide some of this identification, for example all river stage observations may be held in a single file, and the times of observations will be given by their ordering within the file. In other cases the place or time of observation will have to be explicitly stated, for example a bore-hole log record would contain the coordinates of the bore-hole. By grouping the data the amount of extra information which needs to be stored can be reduced. In the subsequent paragraph it is discussed that how data can be broken down into groups for storage, and defines those groups and identifies the attributes of each which govern data storage. Also described are the general characteristics of data storage systems and the advantages and methods of coding data.

4.1 Data Types

To assist in identifying data types, an immediate distinction may be drawn between data which is fixed and data which varies in either time or space. In the case of fixed data only a single entry need be made along the axis of the dimension in which the parameter is fixed. Thus, data which is related to a physical object at a specific location, or to a parameter which is constant in space need to be represented by only one entry on the space axis. Similarly, parameters which are fixed in time need only one entry on the time axis. Conversely, variable parameters must be represented by a series of entries along the axis of the varying dimension. Figure 3 presents the possible combination of fixed and variables data in space & time dimensions and gives examples of each type in terms of land and water data.

Time series data may be divided into three groups which relate to the intervals at which observations are made. There are regular time interval, irregular time interval and continuous observations. The distinction is drawn as it affects storage formats. Spatially varied data may be divided into two groups to highlight the difference between data which are sampled at points and requires spatial interpolation, and data that are sampled on an areal basis by remote sensing techniques.

Remote sensed data in the form of photographs need to be manually or automatically digitized for data entry. If the data are satellite based they are probably provided directly in computer compatible form. However, such data usually need compressing before they can be included in the normal scale of data base. Similar problems arise with data sampled continuously with time as these also need to be required, aerial data being reduced to sets of x-y distance co-ordinates, and continuous time data being digitized into an interval time series. Thus, referring to figure 3 computer storage formats are only required for non-time series data types 1 and 2, and time series data types 4, 5, 7, 8.

4.2 Computer Storage Systems

Computer storage systems are almost analogous to the manual system. The storage media is not paper or card but disk or tape. Related data items are collected into logical records, the space reserved for each data item being called a field. Data items which will be used as a basis for record retrieval are called keys. The most important key is the major record identifier. The units of storage media on which records are written are physical records rather then sheets of paper. Several logical records may occupy a physical record or vice-versa. Similar records are grouped together into files and several files are normally held in one physical storage device. In the case of computer this is a tape, disk pack or diskette, also called a storage volume. Very large files may occupy more then one storage volume (multi-volume files as opposed to multi-file volumes). Each storage volume has a master file or catalogue containing the location and structure of the files held.

When data are retrieved by passing records one by one until the required record in found, this is called serial or sequential accessing. The alternative approach of creating indexes for each desired key and passing directly to any required record is called direct or random accessing. Files which are permanently accessible need to be held on mass storage devices, almost invariably disks, and are said to be on-line. Files not held on the machine are said to be off-line or achieved. One feature of conventional storage systems used extensively in computer systems is the coding of data to reduce the bulk and improve the intelligibility of large volumes of data.

4.3 Coding of Data

Whatever type of data system we have, we must be able to identify each data item clearly, to classify it in connection with other similar data items, and to sort, retrieve, or otherwise process it by means of machines. In data processing, then, we make widespread use of codes, and a good portion of the system analyst's work is to develop, maintain, and analyse codes of various types.

A code may be defined as a brief little, composed of either letters or numbers, used to identify an item of data and to express its relationship to other items of the same or similar nature. A good code should serve most or all of the following objectives:

a) Uniqueness : e.g., a large file of names of villages (stations) may contain some of villages with the same name. A unique code number positively identifies the village (station) in question.

b) Brevity e.g., The codes WV and PH are shorter than the descriptions wind velocity and historical precipitation. Other may show an even more dramatic reduction in length.

c) Speed : Most machines are able to process coded data much more rapidly then they can handle the original descriptive information.

d) Sequencing : Codes may be assigned in such a way that data may be placed in any desired sequence. Codes may thus be assigned according to geographical area, frequency of use, cost, or any other desired sequence that bears no relationship to alphabetic order.

e) Uniformity : Codes make it possible to identify similar items, even though terminology used to describe them may differ.

5.0 DATA BASE DESIGN

Data base is the nucleus of the data-processing system. Data base represents the information "asset" of any organization. It is possible to conceive a storage system where each data application has its own data file, or at least its own group of data files containing the necessary time series and non time series data. This would be a very efficient way of storing data for that particular application, and other files could similarly be created for each other analysis. However, it is not difficult to see that this is extremely inefficient in terms of computer storage because of the duplication of data items. Many files would contain identical values of rainfall, river flows, soil type, etc. Even if storage were not limited there would be problems in keeping many copies of the same information up to date.

At the opposite end of the scale each data item could be stored only once, subsequently bringing together those data items required for each application. This eliminates completely any duplication of data, but causes some problems in relating data items for every application. In computer terminology this is an "integrated data base".

To eliminate the duplication of data in this way, the data must be structured and organized to include the requirements of all users. Data should not be recorded on local files designed for specific applications. The concept is simple, and hardware and software capable of managing very large data bases are now available. These software systems are called data base management systems (DBMS).

5.1 The Role of DBMS

Except for the special case of multiple parameter series, it has been the practice with time series data to use only single parameter files. Thus, separate files are kept for rainfall, river flow and groundwater levels etc. Problems mostly arise when seeking to group non time series data and it was largely for this reason that DBMS software was developed. A DBMS effectively acts as an interface between the data base and the data application, and provides a method for defining all the necessary data relationship for each application. To perform this task a data description language (DDL) is used.

Whilst a DBMS can easily be specified for most computers, there are several reasons for being cautions in their application given the lack of computing experience. The main reasons are:

a) DBMS requires a very high level of technical skill to implement and maintain;

b) If data has been structured by one DBMS which is subsequently found inadequate or inappropriate, the task of converting to a new system, or modifying the old one can be very substantial;

c) The options regarding record size, formats and relationships are frequently too inflexible.

6.0 OPTIONAL ASPECTS OF DATA PROCESSING

Once established, the data base system operation is a continuous cycle of data input, verification, updating and retrieval. Methods described subsequently have to be employed for maintaining the quality of data being stored and its subsequent security.

6.1 Accuracy of Keyed Data

Data entry is most commonly carried out using key punches. Key punching can be performed on line, when data is keyed directly onto the computer mass storage device under computer control, or may be off line when data is keyed onto some intermediate computer compatible media such as cards, tapes or diskettes. However, the functional features of any physical system for entering the data onto the computer storage are the same. The problems encountered are ensuring the accuracy of data coding and key punching. The usual method of ensuring that data has been key punched correctly is to have it punched twice. This is the technique known as punching and verifying and is described under section of data processing (in 2.0).

The next stage is to enter the data into the computer and check that it is within the correct range of values for numeric data, and uses allowable codes for coded data. This validation provides some check on the accuracy of coding and to some extent key punching, although the ordinary key punching and verifying should have removed most errors from this source. Validation routines may be written by the user or provided as a part of a comprehensive DBMS. During this process the data may also be converted into the final form in which it is to be stored in the data base. This may require the aggregation or splitting of input records to produce records in the required storage format.

Coded data are best checked by preparing decoded reports of the data in a form as close to that of the original data prior to coding. A comparison of the two data sets can then be made.

6.2 Accuracy of Chart Data

Ensuring the accuracy of chart or map data is a more complex process. After entering data from a chart, the most simple means of verification is a direct plot or VDU display of the input. Once this initial verification has taken place, the digitized chart must be corrected for time, calibration and amount.

6.3 Validation and Updating of Data

When data have been key punched or otherwise converted to computer compatible form they are entered into the data base. The records of data will be stored in some order based upon sorts using one or several keys. The result may be a file sorted such that the key values are put into a numerically ascending, or perhaps alphabetical order. When creating a new file the data are validated on entry and all validated records are saved and sorted. The records requiring correction are key punched and put with the next batch of data to be added to that file. The new batch of data is read in and validated, the validated records are saved on an intermediate file and sorted. Finally, the data on this file are merged with the data on the original file. The merging process ensures that the data are maintained in the order of the key(s). The process of entering corrections can be kept separate from normal augmenting or updating of information if required, though the operations are the same Fig. 4 helps to explain this process.

7.0 DATA RETRIEVAL

Previous sections of this publication have described the preparation, input and storage of data. Whereas these are the operational methods, a reliable and comprehensive data retrieval system is one of the principle objectives of a data processing system.

There are generally two levels of retrieval data. The first is to abstract blocks of related data from one or several files in the data base using some criteria. These blocks themselves become new files which may be used as an input to some subsequent analysis. The second level, made attractive by the development of on line user terminals, is the direct interrogation of the data base and the selection, display and analysis of data. The first task is becoming increasingly absorbed into the capabilities of the second.

An overriding consideration with all retrieval systems is that the same retrieval program can be used as wide a range of storage records as possible. This is greatly facilitated by the use of common storage formats and of header records in each file which inform the retrieval program of the record structure. A retrieval system formulated in this way is said to be "data base independent", that is, the retrieval program does not know, or care, what are the actual data being manipulated.

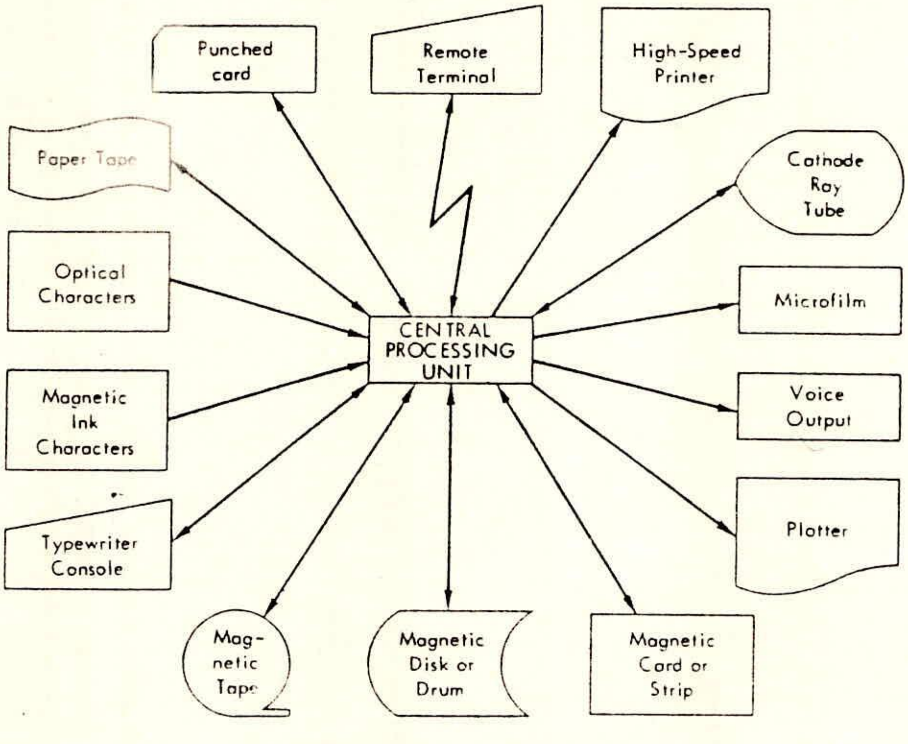


Fig. 1 Typical hardware of a modern computer system.

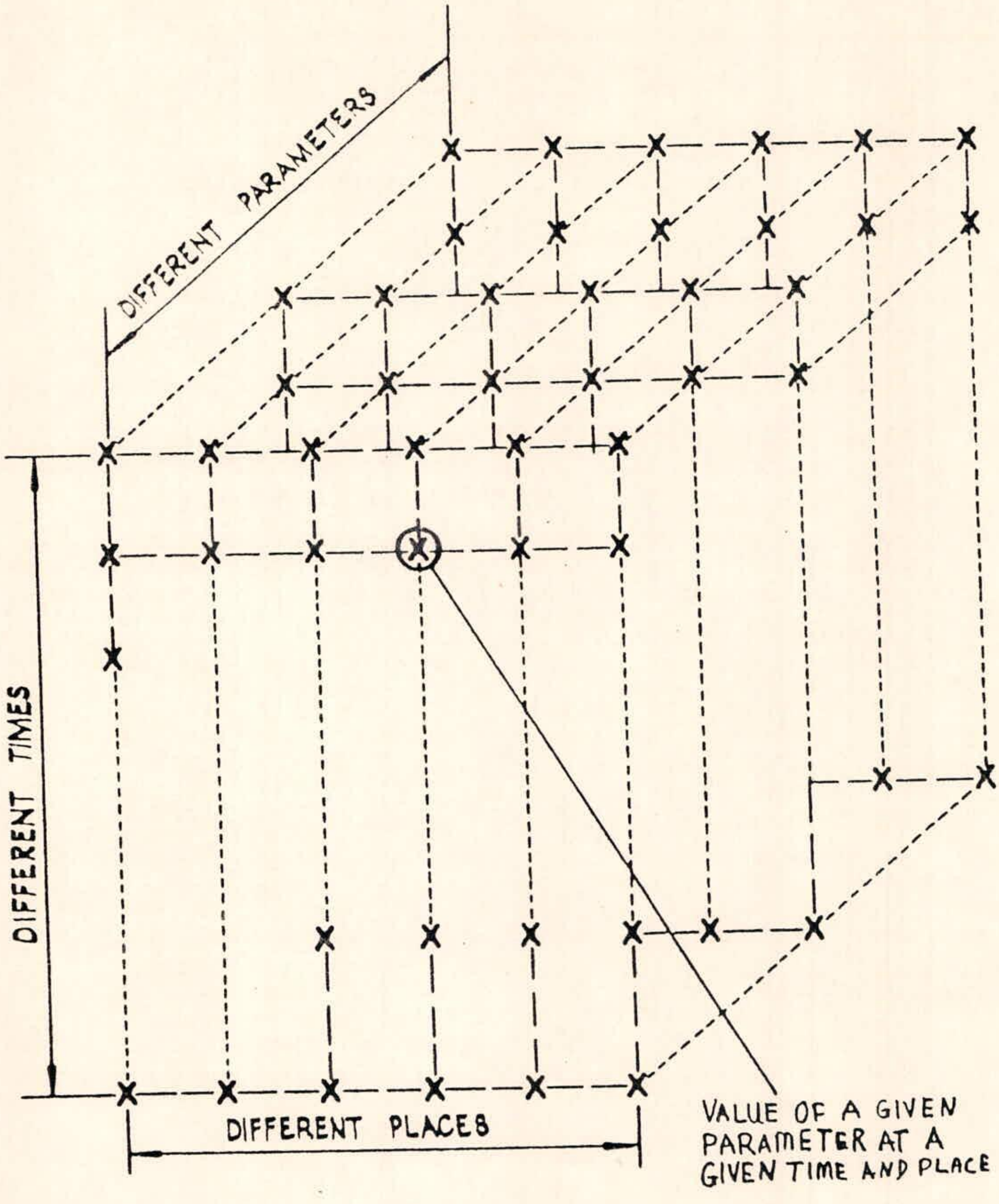


Figure 2 - The three-dimensional nature of resources data

			SPACE		
			FIXED PARAMETER	VARIABLE PARAMETER	
				POINT SAMPLING	AREA SAMPLING+
TIME	FIXED PARAMETER (Non-Time Series Data)		1. O-O-N Physical works: Dams, pumps, wells, boreholes etc.	2. O-R-S Soil profile data.	3. O-C-D Topographical data from aerial surveys.
	VARIABLE PARAMETER (Time Series Data)	REGULAR SAMPLING	4. R-O-T Monthly reservoir levels.	5. R-R-TS Daily rainfall data.	6. R-C-D Land use by orbiting satellite.
		IRREGULAR SAMPLING	7. R-O-T Gauging of ephemeral springs.	8. R-R-TS River flood gauging.	9. R-C-D Land use by aerial photography.
		CONTINUOUS SAMPLING	10. C-O-D Chart recording of reservoir discharge.	11. C-R-D Chart recording of borehole water level.	12. C-C-D Meteorological data from geostationary satellite.

Notes: 1. Sampling Code: X-Y-Z where X and Y are space and time sampling frequencies, Z is a data interpretation code.
Codes for X and Y values: O - Sampled once only. R - Sampled repeatedly. C - Sampled continuously.
Codes for Z values: T - Needs interpolation in time. S - Needs interpolation in space.
N - Needs no interpolation. D - Needs digitising.

2. + Satellite data may be in digital form but will need compressing.

Figure 3 - Space-time combinations of data

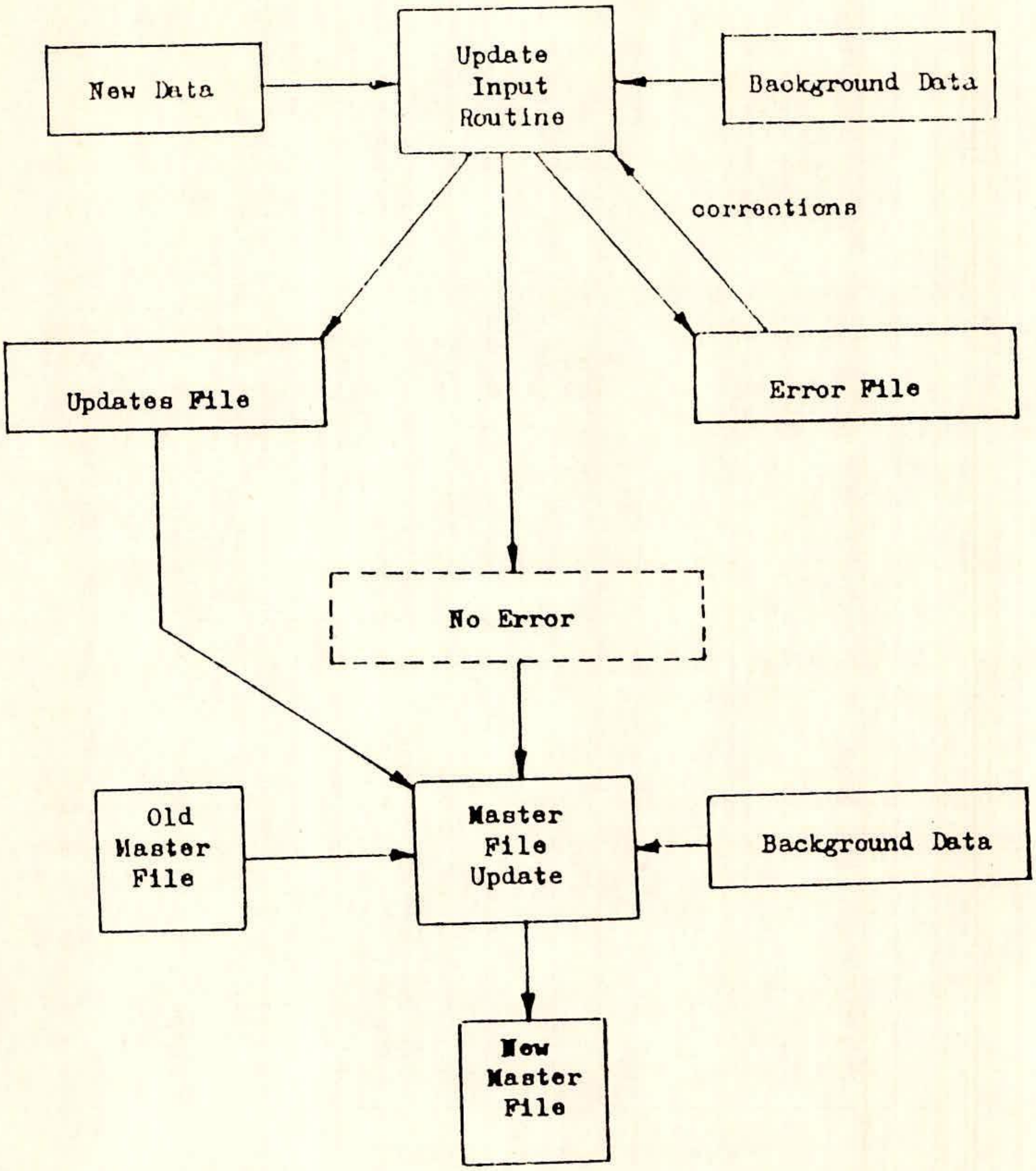


Figure 4 - File updating logic

TABLE 1

The components of data collection

DATA COLLECTION		
DATA CAPTURE		TRANSMISSION
SENSING	RECORDING	
1. Visual Water level gauge, land use, site description, soil texture etc.	1. Field note book Text descriptions and element or parameter values.	1. Manual Field observers Postal service Telephone
2. Mechanical Raingauge, thermometer, current meter, soil penetrometer.	2. Field data sheet Purpose designed for particular text descriptions and element or parameter values. May be pre-coded for subsequent computer input purposes	2. Automatic (Telemetry) Telephone Dedicated land line Radio Satellite
3. Electrical Thermistor, radiometer, pressure transducer, conductivity probe.	3. Charts Strip charts with element value continuously recorded by pen tracing	
	4. Computer compatible media (a) Manually recorded Mark sense forms Multiple choice forms (b) Automatically recorded Cassette Paper tape Solid state memory	

NOTE: The table applies to elements or parameters observed in the field. There are notable groups of data, e.g. in soils and water quality, where laboratory analysis of physical samples are performed. Here the data-collection system will almost invariably be:

(a) Mechanical sampling;
(b) Note book/data sheet field entries;
(c) Transport of samples to laboratory by field staff.

TABLE 2

The components of data processing

DATA PREPARATION	DATA ENTRY (INPUT)	VALIDATION	PRIMARY PROCESSING	DATA BASE UPDATING	SECONDARY PROCESSING	RETRIEVAL	OUTPUT
			DATA PROCESSING				
Prepare punching documents by: 1. Transcription Field note book entries Non-standard data formats 2. Coding Reduction/standardisation of input data	1. Punching document a. Direct keying through VDU b. Keying into computer compatible media Punched cards Key-to-tape Key-to-disk 2. Charts and maps Direct input by digitiser 3. Computer compatible media a. Tapes/cassette b. Diskettes c. Cards d. Communication lines (telemetred data) e. Mark sense/optical character readers	1. Range checks 2. Sum checks 3. Inter-station consistency checks	1. Standardisation of units 2. Calculation of derived parameters 3. Further coding of input to reduce storage requirements 4. Arranging data in data base format	1. Add new data sets to existing data base 2. Report any errors	1. Programs for routine reports 2. Statistical summaries 3. Infilling missing data values 4. Interpolation or aggregation of data	1. Selection of data by: a. Parameter type b. Parameter value c. Location d. Period of record e. Time interval of record 2. Selection of output device	1. Printers 2. Plotters 3. VDU 4. Computer storage media 5. Microfilm 6. Telemetry
	ERROR CORRECTION						

PART II

A SYSTEM FOR STORAGE PROCESSING AND RETRIEVAL OF HYDRO-METEOROLOGICAL DATA (HYMOS)

PART II

A SYSTEM FOR STORAGE, PROCESSING AND RETRIEVAL OF HYDRO-METEOROLOGICAL DATA (HYMOS)

1.0 GENERAL OVERVIEW

HYMOS is a database management and processing system for hydrometeorological quality and quantity data, designed for use on personal computers (pc's). It arranges a convenient structuring of data in a database and provides an extensive set of tools for data entry, validation, completion, analysis, retrieval and reporting. It is comprehensive, well tuned and easy to use via full screen menus with on-line help to guide the user. The package includes many tabular and graphical options facilitating efficient reporting. It runs on stand alone computers, but can also be used in a network system. Securities have been built in to restrict the access for certain activities to qualified staff only.

HYMOS is developed to streamline the storage and processing of (geo-) hydrological and meteorological data. It is tailored for use by hydrological and meteorological data processing branches, water resources management authorities, water boards, water engineering consultants and hydrological advisers. HYMOS data are to a large extent typically time-oriented. Together with a space-oriented Geographical Information System, it covers all data storage and processing requirements for planning, design and operation of water management systems.

In this introduction an overall view of HYMOS is given, the type of data it can handle, the way the data can be stored and processed and the hardware required to run the system successfully.

1.1 Computer Configuration Requirements

The HYMOS software package is written in the programming languages FORTRAN 77 and C.

The HYMOS version 3.0 runs on computer systems of the following specifications:

. IBM XT, AT or compatible personal computer, with:
  . 640 kb core memory,
  . ≥ 20 Mb hard disk,
  . EGA or VGA graphics card,
  . MS-DOS 3.* (or higher) operating system,

. Printer
. HP or Calcomp compatible plotter;
. tape streamer (recommended for back-up purposes):
. protection key HASP II (Supplied by the developer)

For fast and pleasant operation an AT computer with color monitor is preferred but not strictly necessary, nor is a co-processor required. Graphics cards less advanced than EGA or VGA do not support all HYMOS graphics, but they will be acceptable for common data processing activities.

1.2 HYMOS in a nutshell

1.2.1 Structure of HYMOS

HYMOS integrates the distinctive phases in the processing of hydrological data. The activities are carried out in specific processing modules; each module consists of a number of programs enclosing particular compilations/computations. The modules are structured according to a logical sequence of activities in data processing. All modules are linked to the HYMOS database, structured and controlled by a database management system. Different databases can be operational under HYMOS. In summary, HYMOS comprises of following systems:

A. a database management system, to create databases, to structure the database and to define user identifications;

B. a data storage and retrieval system, covering data entry, editing, reporting in tabular and graphical form as well as the transfer and retrieval of data;

C. a data processing system, including validation, series completion by interpolation, simulation and regression techniques, elaboration of flow measurements, data compilation, statistical analysis, and time series analysis.

The structure of HYMOS is shown schematically in Figure 1.

HYMOS allows user programs to communicate with the database. For that purpose the FORTRAN subroutine library HYSUB is enclosed with the package. HYSUB comprises an extensive set of subroutines to store and retrieve data and to carry out statistical analysis.

1.2.2 Data types

The types of data, handled by HYMOS, can be categorized in the following groups.

1. Space-oriented data, covering:	-catchment characteristics -station particulars -station histories -geo-hydrological profiles
2. Time-oriented data, including:	-equidistant time series, i.e. series with regular time intervals -non-equidistant time series, i.e. series with irregular time intervals.
3. Relation-oriented data, divided in:	-stage-discharge data -relation or rating curves parameters, valid for a certain time period.

1.2.3 Database management

Database management deals with the creation and structuring of data bases and with the definition of user identifications.

Structure of databases:

A HYMOS created database comprises the hydro-meteorological data of one or more catchments, or sub-catchment(s). As many databases as required can be used. Each database is stored in a separate directory.

In a particular database the data are, except for the catchment characteristics, structured station-wise. Data type and time interval are used for further identification of station data. The overall structure is shown in the following scheme:

```
                                          Station particulars
                                          Station history
                       Station a          Geo-hydrological profiles
    Catchment j ----   Station b   ----   Equidistant time series
                       Station c          Non-equi. time series
                                          Stage discharge data
                                          relation curve parameter
                                          rating curve parameter
```

Definition of users:

User identifications can be categorized into three parts:

1. group name,
2. user name, and
3. password.

The access to a particular database is reserved to specified groups. Within a group users are defined with different levels of authority, ranging from access to all facilities to data retrieval operations only. Each user must enter a unique password, which can only be changed by the System Manager.

1.2.4 Data storage and retrieval

Under HYMOS the data are stored in dBase look alike files (space-oriented data, non-equidistant time series and relation-oriented data) and in dedicated direct-access files (equidistant time series). Hardly any limits exist to the size of a database as may be seen under section of 'Data Limitations'.

To store data in, and to retrieve data from the HYMOS database, the following modules are available:

* data entry and editing:

  HYMOS provides three ways to load the data base, viz.:
  - via data files on diskette or in EPROM,
  - manually, via the screen, and
  - via one of the HYMOS processing options, in case of

computed data.

Full screen editors are available under HYMOS for all data types to add, edit, display and delete data.

Codes are generated by HYMOS and stored in the data base to distinguish among original, corrected and filled-in data.

* reporting and retrieval:

Various entries can be used to retrieve data from the database, the for purpose of transfer to ASCII files or other databases either or not under HYMOS. Ready made monthly or annual reports can be produced by HYMOS; many tabular options and powerful graphics are available to support the reporting.

1.2.5 Data processing

The validation, completion and analysis of hydro-meteorological data under HYMOS is logically structured in the following data processing modules:

* data validation:

tabular, graphical and computational procedures are available for proper screening of various types of data.

* data completion and regression:

a number of time and spatial interpolation techniques, as well as powerful regression and rainfall-runoff simulation (Sacramento model) are included for series completion.

* flow measurement:

procedures are provided for elaboration and checking of current metering data, stage discharge analysis and conversion of stages into discharges.

* data compilation:

including aggregation and dis-aggregation of series, series transformation, computation of average and extreme values, catchment rainfall and evapotranspiration computation.

* statistical analysis:

computation of basic statistics, fitting of distribution functions, statistical tables, random data generation, computation of IDF-curves and frequency and duration curves.

* time series analysis:

covering correlogram and spectral analysis, range and run analysis and computation of storage requirements.

1.2.6 Special features

To facilitate the use of HYMOS special function keys are available for on-line help, display of temporary output files, use of text editors, change of data base and packing of data base files.

The station and series selection is enhanced by selection from displayed tables or catchment maps on the screen.

Finally, keys are available to set the screen colours and hardware configuration, and to switch to DOS during a HYMOS session.

1.2.7 Database limitations

It was mentioned earlier that the equidistant time series are stored in special direct access files. Some limitations exist as to the amount of data, the number of series and the length of the series that can be stored in these files. The following limitations apply to a particular data base.

. the amount of data is at maximum 1.8 billion (i.e. 7 Gigabyte);
. the amount of equidistant time series is at maximum 10,000;
. all series can have a length of at least 54,500 data (e.g. about 150 years of daily data) but, in addition, some 40 million data positions are available for extension of one or a number of series.

The last observation allows one to use for a limited number of series a length far beyond the standard of 54,500 data, e.g. 25 series of 1.6 million data (i.e. 45 years of 15 minute data) can be used additionally. For non-equidistant and non time series data no limits exist, but the size of the disk.

If the above limits are exceeded, more databases have to be defined, e.g. a temporary data base with a few permanent data bases, where the latter are categorized according to location, data type or time interval, depending on the objectives of the Information System.

2.0 TYPES OF DATA

2.1 Overview

The type of data that can be stored in HYMOS include:

1. Space oriented data:
   - -catchment characteristics,
   - -station data: characteristics, log-book and histories,
   - -series characteristics, and
   - -geo-hydrological profiles.

2. Time oriented data:
   - -equidistant time series, and
   - -non-equidistant time series.

3. Relation oriented data:
   - -stage-discharge data
   - -rating curve parameters, and
   - -relation curve parameters.

The contents of the various types of data are presented in the sections from 2.2 to 2.8.

2.2 Catchment data

2.2.1 General

Catchment data comprise:

1. latitude and longitude of local origin of catchment data,
2. catchment layout data, and
3. (sub-) catchment characteristics.

2.2.2 Local origin

The local origin represents the lower left corner of the catchment map and is given in geographical coordinates. Catchment layout-data, see section 2.2.3, are entered in kilometers east (x-coordinate) and north (y-coordinate) of the local origin. There is only one local origin for a particular data-base, valid for all catchments or sub-catchments in the database!

2.2.3 Catchment layout

Catchment layout data includes:

- -catchment boundaries,
- -plan form of river(s),
- -location of cities,
- -network of roads, and
- -catchment topography

2.2.4 Catchment characteristics

In a database, data of one or more catchments or sub-catchments can be stored. For each catchment or sub-catchment this comprises the following characteristics:

- -(sub-)catchment name,
- -name of the river,
- -(sub-)catchment area
- -stream order,
- -length of the river, and
- -slope of the river.

2.3 Station data

2.3.1 General

Station data covers:

1. station code,
2. general station data,
3. station log-book,
4. station history, and
5. series characteristics.

2.3.2 Station code

The station code is a unique set of 1 to 10 characters (letters, digits or symbols) to identify a station. The easiest way of coding a station is by using the station name or an abbreviation of it.

Example

For Automatic Hydrological Station (AHS) at NIH, Roorkee (30^{o}15'00", 78^{o}30'00") in Uttar Pradesh, India, e.g., the following station codes could be used by HYMOS.

ROOR-NIH
NIH-AHS
3015 7830
etc.

2.3.3 General station data

General station data include:

- station name,
- name of the river (for streamflow stations),
- district and country,
- station latitude, longitude and altitude,
- catchment area upstream of the station, and
- agency, responsible for the station.

2.3.4 Station log book

The log book contains user remarks on series of the station. These remarks typically refer to processing instructions. The information stored includes:

- the remarks,
- period, the remarks refers to,
- series it concerns,
- status of execution of the remark,
- additionally required action,
- date of the remark, and
- name of the person who made the remark.

2.3.5 Station history

The station history is a text file with user specified information about the station. There is no restriction on the layout or contents of the history file. Typically, the station history comprises information on the establishment of the station, access to the station, benchmarks, gauge particulars, repairs, etc.

2.3.6 Series characteristics

Series characteristics refers to data type and time interval parameters, which are required to organize the storage and retrieval of data from the database.

With respect to series a distinction is made between:

1. equidistant time series,
2. non-equidistant time series, and
3. special time series.

The equidistant time series characteristics include:

- number of the data file (= HIS - file) in which the series is (to be) stored, see section 3.2.
- data type, unit and type of observation,
- time interval,
- basic time interval,
- time shift,
- missing value,
- likely minimum and maximum value, and
- likely maximum rate of rise and of fall (when relevant).

The equidistant time series characteristics are discussed in detail in section 2.5.

The non-equidistant time series characteristics include, see also section 2.6:

- data type, unit and type of observation,
- missing value,
- likely minimum and maximum value, and
- likely maximum rate of rise and fall (when relevant).

The special time series, refer to monthly maximum rainfall amounts and annual exceedances of fixed durations. These series are exclusively applied for Intensity-Duration-Frequency analysis, see section 9.5 .

The characteristics of special time series include:
- fixed data types and unit, and
- missing value.

Basically, these special time series are stored as ordinary equidistant time series, but a different interpretation is given to the time interval.

2.4 Geo-hydrological profiles

2.4.1 General

The geo-hydrological station data comprise information on filters and layers. The overall data include:

-ground level (GL)	(m+MSL)
-number of filters, and	
-number of layers.	

2.4.2 Filter data

The filter data comprise:

-measuring height	(m+GL)
-upper side of filter	(m+GL)
-lower side of filter	(m+GL)
-diameter of well, and	(m)
-length of well + filter	(m)

2.4.3 Layer data

For each layer the following data are stored:

-depth	(m+GL)
-thickness	(m)
-layer code,	-
-geological information,	-
-type of material,	-
-availability of sieve analysis	(Y/N)
-porosity	(%)
-specific yield/storage coeff.	-
-hor. and vert. permeability	(m/day)

2.5 Equidistant time series

2.5.1 Definition

A time series is a collection of data, ordered sequentially in time, representing the behaviour of a process at a fixed position in space.

An equidistant time series is a sequence of data with a fixed time interval. A time series may be divided into sub-series. Sub-series are not necessarily sequential in time. Time gaps of any size may be present between the subseries.

To specify equidistant time series (series code) following elements are used in HYMOS.

-station code,
-data type, and
-time interval

Additionally, following characteristics are stored

-data unit,
-type of observation,
-basic time interval,
-time shift,
-missing value,
-data limits.

The time labels of the data are derived from the time interval and are based on calendar and time units.

2.5.2 Data type

The data type is a two character specification of the type of data. Any combination of two characters (if not used for non-equidistant time series and not equal to AE or AM) is permitted and can be added to the available set.

EXAMPLE

PH	-	historical precipitation
PG	-	generated precipitation
H1	-	water levels upsteam of structure
H2	-	water levels downstream of structure
QH	-	historical discharge

To each data type belongs a data unit and a type of observation, (e.g. accumulative or instantaneous).

2.5.3 Units

The data unit is a characteristics of the data type. For a specific data type one and only one unit applies. Any unit can be used for the data in HYMOS, however, the use of SI-units is strongly recommended.

2.5.4 Type of observation

The type of observation is another characteristic of the data type, and it refers to the sampling or computational procedure used to obtain the data. Following types of observations are distinguished for equidistant time series, see also Figure 2.

- -instantaneous observations,
- -accumulative observations, and
- -average observations

Instantaneous observations:

The observation is made at a certain point in time within the time interval Δt; it produces one instantaneous exposure of the process in a time span Δt, like e.g. water level time series, discharge time series.

Accumulative observations:

The observation is the integral of the continuous process in the time interval t to $t+\Delta t$, like for example rainfall amounts. Accumulative observations produce volumes rather than intensities. For example runoff expressed in mm is accumulative, whereas runoff expressed in m^3/s is instantaneous.

Average observations:

Average observations are computed quantities, obtained e.g. by aggregation of instantaneous observations over a larger time interval. Average observations are treated in HYMOS as instantaneous observations since in common mathematical operations they are equivalent. (From a view-point of information content they are, however, different: by averaging, information about the process is lost!).

The type of observation has important consequences for aggregation of time series (i.e. averaging versus summing) as well as for dis-aggregation of time series (see section 8.1).

2.5.5 Time and calendar units

To define time intervals between and times of occurrence of series elements use is made of the following time and calendar units in HYMOS:

* time units	-minutes
	-hour.
* calendar units	-day
	-month
	-year

The Calendar units vary with time according to the Gregorian calendar. This implies:

1. months of 28, 29, 30 or 31 days; January 31 days, February 28 or 29 days, March 31 days, April 30 days ,etc.

2. non-leap year:28 days in February and 365 days in a year,
   leap year :29 days in February and 366 days in a year.
   a Leap year is defined as follows:
   - for non-secular years : a year divisible by 4,
   - for secular years : a year divisible by 400.

Secular years are the years at the turn of the centuries, e.g. 1800, 1900, 2000,2100. So, the year 2000 is a leap year, but 1900 is not.

2.5.6 The time interval

The time interval Δt represents:
- the time distance between successive series elements of instantaneous observations, or
- the time span of an accumulative observation

The time interval may be year, month, day or hour or a part thereof. Generally ,

Δt = (time interval unit, divider)

where:

time interval unit = year, month, day or hour (i.e. a calendar or a time unit), indicated by an integer:

Year = 1,	day = 3
month = 2,	Hour = 4

divider = Division factor (integer : 1-99) applied to the relevant time interval unit.

This definition provides a very flexible method of specifying time intervals. Practically all commonly used intervals between one minute and one year fit in the definition.

EXAMPLES

To demonstrate the use and flexibility of the time interval inspect the following .

Interval	HYMOS	Explanation	Interval	HYMOS	Explanation
(Year,1)	(1,1)	Year	(day,1)	(3,1)	Day
(Year,4)	(1,4)	season	(day,3)	(3,3)	8 hours
(month,1)	(2,1)	month	(hour,1)	(4,1)	hour
(month,3)	(2,3)	decade	(hour,4)	(4,4)	quarter
(year,52)	(1,52)	week	(hour,60)	(4,60)	minute

NOTES

1. For some intervals there is no unique combination of unit and divider, e.g. month may be specified by (2,1) but also by (1,12): in such cases preference is to be given to the definition with the lowest divider.

2. Some combinations of time interval units and dividers lead to non-equal time intervals (e.g.decade and week).
   -decade :if the number of days in a month ≠ 30, then one of the decades ≠ 10 days
   -week : a year is not exactly divisible by 52, hence one of the weeks ≠ 7 days, etc.

   To define irregularity, the basic time interval is introduced. It is, in practice, only used when the time interval is a week, a pentad (i.e. a period of 5 days) or a decade. The definition of the basic time interval is explained in the next sub-Chapter.

2.5.7 Basic time interval

In case the user cannot or does not want to divide the time interval unit into equal parts then in addition to the time interval Δt, a basic time interval or basic Δt expressed in numbers of smaller calendar or time units has to be applied. E.g. for decades, which are parts of months, multiples of days are used to specify the interval; similarly for weeks, which are parts of years, multiples of days are used to specify the length of the interval.

The basic Δt is defined as follows:

Basic Δt = (basic time interval unit, replicator)

where:

basic time interval unit -	month, day, hour or minute (calendar or time unit) indicated by an integer: month = 2 hour = 4, day = 3, minute = 5.
replicator -	multiplication factor applied to the basic time interval unit.

Consider e.g. decade intervals, defined as Δt = (month, 3) = (2,3). To specify that a decade consists of 10 days the basic time interval should read: basic Δt = (day,10) = (3,10). Then as much as possible intervals of 10 days will be defined in a month. So the first 2 decades will contain 10 days, while the last decade comprises 8, 9, 10 or 11 days depending on month and year.

This leads to the following generalization. The time interval Δt is built up out of a number of basic Δt's equal to the divider. The last time interval within a time interval unit may contain an amount of basic time interval units, that differs from the replicator. Let the time interval unit, containing n basic time interval units, be divided into K parts and let the replicator in the basic Δt be m. Then the first k-1 time intervals comprise m basic Δt units. The last or K-th time interval contains (n-(k-1).m) basic Δt units, see also Figure 3.

EXAMPLES

Pentad intervals: Δt = (month,6) = (2,6) and basic Δt = (day,5) = (3,5). The first 5 pentads of the month contain 5 days, while the last pentad includes 3, 4, 5, or 6 days depending on the month and year.

Weekly intervals: Δt = (year,52) = (1,52) and basic Δt = (day,7) = (3,7). The first 51 weeks of the year contain 7 days, while the 52nd week includes 8 or 9 days.

2.5.8 Time label

Based on the definition of the time interval Δt, the positioning of a series element in time, i.e. the time label, is expressed in calendar and time units and a subdivision index. The subdivision index determines the position of the series element within the applied time interval unit; its value can be at maximum equal to the divider.

Generally, the time label or date of a series element is given by:

t = (year, month, day, hour, subdivision index)

where :

year	:	>1850
month	:	1-12
day	:	1-28, 29, 30, 31
hour	:	0-23
si	:	1-99 [(s)ub division (i)ndex]

NOTE

The elements of the time label refer to positions in time and not to time intervals; they are only expressed in multiples of time interval units. Only those elements of the time label have to be used which are relevant to position the series elements; if an element of the time label is not in use a zero is entered.

EXAMPLES

Date		HYMOS Time label
30 January 1989	-	1989, 1, 30, 0, 1
Decade of 11-20 February 1976	-	1976, 2, 0, 0, 2
Quarter 24.00 28/2/1960-00.15 29/2/1960	-	1960, 2, 29, 0, 1
12th week of 1991	-	1991, 0, 0, 0, 12

2.5.9 Time shift

The quantity Δt time shift is introduced to define the position of an observation inside the time interval Δt more accurately, e.g. for a daily rainfall series one wants to specify that the gauge is daily read at 8.00 hrs. This quantity should be considered as a property of a series and as such it has no meaning for the positioning of a series element in time in the database. In the set-up of the database this time shift does not play a role. Only when at a later stage a number of series are mutually compared, the information on a time shift within the time interval may be of importance; for example, daily rainfall totals derived from quarterly or hourly observations often refer to the period 0.00 - 24.00 hrs, whereas daily rainfall totals from non-recording gauges generally refer to days from e.g. 8.00 to 8.00 hrs. the next day and hence are not mutually comparable.

The shift inside Δt (positive only) is expressed in lower order calendar and/or time units with an accuracy not more than 1 hour.

The Δt time shift has 3 elements:

Δt time shift = (number of months, number of days, number of hours)

EXAMPLE

Daily rainfall, measured at 8.00 hours:

Δt	=	(day,1)	=	(3,1)
Δt time shift	=	(0,0,8 hours)	=	(0,0,8)

2.5.10 Missing data

For each series a value is stored in the database to indicate that a data point is missing. Although the choice of this value is free it should always be less than the lower possible value.

EXAMPLE

Typical values for missing data are:

Rainfall	-1.	
Temperature	-99.	
Water level	-1.	provided that no negative levels occur (depends on the zero-level of the gauge)
Discharge	-1.	in non-tidal areas

-9999. in tidal areas, provided that the flow is always > -9999.

2.5.11 Data limits

For validation purposes the likely minimum and maximum values of a series as well as the likely maximum rate of rise and of fall are stored as series characteristics. In some validation options the series elements, which exceed these limits, will be flagged.

2.5.12 Series code

The series code is formed by:

-station code	:	1 to 10 characters
-data type	:	2 characters
-time interval	:	2 numbers: a time or calendar unit (1-4) and a divider (1-99).

So the total series code occupies 15 positions. To store and retrieve equidistant time series data the full series code must be applied. Special function keys are available to ease the selection.

EXAMPLES

Code of rainfall, water level and discharge time series of the hydro-meteorological station with station code ROORNIH could read:

ROORNIH	PH3 1	=	historical daily rainfall
ROORNIH	PG1 1	=	generated annual rainfall
ROORNIH	HH4 4	=	historical 15 minute water levels
ROORNIH	QG152	=	generated weekly discharges.

2.6 Non-equidistant time series

2.6.1 Definition

Non equidistant time series are sequences of data unequally spaced in time. This implies that each data point has to be stored with the time label. To specify these type of series following elements are used in HYMOS:

-station code, and
-data type

The station code has been described in section 2.3.2.

2.6.2 Data type

The data type is a two character specification of the type of data. Any two character combination (if not used for equidistant time series and not equal to AE or AM) is permitted and can be added to the available set.

To each data type belongs a data unit and a type of observation, (e.g. instantaneous or constant).

2.6.3 Units

The data unit is a characteristics of the data type. For a specific data type one and only one unit applies. Any unit can be used for the data in HYMOS, however, the use of SI-units is strongly recommended.

2.6.4 Type of observation

Sampling of continuous processes can be done in several ways, leading to different types of observations. For non-equidistant time series following types of observations are relevant:

- instantaneous observations
- constants

Instantaneous observations :

The observation is made at a certain point in time. It produces one instantaneous exposure of the process, e.g. a groundwater level observation. In between two observations the value of the process is uncertain.

Constants :

The process remains constant from one observation until the next. This is typically representative for time series with human interference, like e.g. gate levels fixed for a period of time or pumps in operation.

2.6.5 Data specification

The time labels of data consist of:

-calendar units : year, month, day
-time units : hour, minute

Only the relevant calendar and time units have to be specified.

2.6.6 Missing data

For each series a value is stored in the database to indicate that a data point is missing. This value is always less than the lowest possible value.

2.6.7 Data limits

For validation purposes (e.g. to flag unlikely data) the likely minimum and maximum values of a series as well as the likely maximum rate of rise and of fall are stored as series characteristics. Series elements exceeding these boundaries are flagged.

2.6.8 Series code

The series code is formed by:

-station code : 1 to 10 characters
-data type : 2 characters

So the total series code occupies 12 positions. To store and retrieve non-equidistant time series data the full series code must be applied.

EXAMPLE

Non-equidistant time series of groundwater levels collected at station with code ROORNIH.

ROORNIH GH = historical groundwater level

2.8 Combined data and parameters

2.8.1 Definition

The combined data and parameters refer to sets of data/parameters of two or more quantities observed irregularly in time or valid for periods of time. These include:

-condensed current metering data,

-discharge and sediment rating parameters, and
-regression/relation curve parameters.

The data comprise observed and computed data/parameters and are entered via the data entry options or created by one of curve fitting options in HYMOS.

2.8.2 Current metering data

Current metering data are condensed results of flow velocity measurements. These data comprise:

-gauge zero	(m above datum)
-water level	(m)
-discharge	(m^3/s)
-gradient or fall	(m/day) or (m)
-river/canal width at the water surface	(m)
-wetted perimeter	(m)
-cross sectional area	(m^2)
-flow velocity	(m/s)
-hydraulic radius	(m)

The original flow velocity measurement data can be entered and stored under HYMOS in dedicated files.

The data are identified by :

-station code
-date (year,month, day) of the measurement
-observation number , and
-type of use

The type of use refers to the application of the flow measurements, either for :

- simple rating analysis, or
- unsteady flow correction (with gradient data), or
- backwater correction (with fall data)

2.8.3 Rating curve parameters

Stage-discharge relations:
Stage discharge relations are described in HYMOS either by :
-a parabolic equation, or
-a power equation.

Where composite cross-sections are present more than one set of coefficients is generally required to describe the stage-discharge relation. HYMOS allows up to 3 sets of coefficients each valid for a specific water level range.

In case of unsteady flow or backwater, a correction is required to arrive at a unique stage-discharge relation. This correction in HYMOS is described as a function of the water level by a parabolic equation and is applied to stages within a certain water level range.

The type of equation, rating curve parameters and the water level ranges are stored in the data base.

The data are identified by :

-station code
-start and end data of the period of validity of the relation

Sediment transport relations:

Sediment transport relations can be described in HYMOS by various equations, e.g. a power equation. The same flexibility exists as for stage-discharge rating curve; also up to 3 ranges on the independent variable can be defined for which parameters can be stored.

2.8.4 Relation curve parameters

Regression and relation curves are described by :
-polynomials
-simple and multiple linear regression equations
-exponential equations
-power equations
-logarithmic equations
-hyperbolic equations

The parameters for these relations with their validity range and period and the series involved can be stored.

The relations are identified by :
-type of equation,
-series combination, and
-validity period.

3.0 Structuring the database

3.1 Database directories

A catchment or sub-catchment forms a separate database under HYMOS. The data of a basin may be stored in one or a number of databases depending on the division into (sub-)catchments.

A particular database is stored in files located in a separate catchment directory,which is a sub directory of the HYMOS databases directory HYMOSDB. The name of the database consists of:

-a 1-4 character database name, prefixed by CAT_,and
-a 1-3 character extension.

Under the catchment directory 3 sub-directories are defined:

-DB	:which stores the database files
-DATA	:which stores particular data files for data processing, and
-FIGURE	:which stores plot files

3.2 Extent of a database

When structuring data into one or more databases, i.e selection of one or more (sub-)catchments, following should be taken into consideration:

1. communication between databases is only possible via series transfer,so
2. all data required for data validation and analysis should be located in to same database, and
3. large database reduce the speed of storage and retrieval operations and require large disks.

The major bulk of the data is generally formed by equidistant time series data. These data are stored in one or more HIS-files located under catchment sub-directory DB. When more HIS-files are used, only the HIS-files, which store the data required for specific data processing activities will have to be loaded. Hence you can save disk space by structuring your data properly into a number of HIS-files. As the HIS-files belong to the same database full exchange of information between the HIS-files is assured.

4.0 Data preparation for transfer

4.1 Introduction

Data can be entered in the HYMOS database in one of the following ways:

1. by reading from ASCII-files
2. via the data entry menus from the keyboard, and
3. through the HYMOS system as a result of computations.

In this section the layout of the ASCII-files for transfer of field data or data from other sources or data bases to the HYMOS database is dealt with for the following types of data:

1. equidistant time series,
2. non-equidistant time series, and
3. stage discharge data.

4.2 Equidistant time series

4.2.1 General

Data are read from data files. Data files may consist of one or more data blocks of one or more series. The way the data are structured in the file and in the blocks is either specified in file- and block-headers or is entered via the screen, leading to the following options:

1. data files with headers, and
2. data files without headers.

For reasons of proper file documentation and safety in data transfer the option with header is strongly advocated. Both options are dealt with in this section.

4.2.2 Data files with headers.

4.2.2.1 File and Block Headers

The data in a particular file may be presented either in free or in fixed format and may be read, see Figure 4:

1. one series per block:time sequentially row-wise, or
2. one series per block:time sequentially column-wise, or
3. multiple series, per block structured parallel, i.e.column-

wise

Further details may vary from block to block and are to be specified for each data block. Hence two types of headers must be used in a data file, viz.:

1. a File Header at the top of a data file, and
2. Block Header(s), one above each data block.

In the File Header the user specifies whether the data is to be read in fixed or free format and whether the data are time sequential row by row or column by column as a single or multiple series. The number of series in the data block, series code(s),start date, data block layout, conversion factor and reference level is contained in the Block Header. This leads to the following general file structure:

Data file :

```
                    File Header
     Block 1:       Header block 1
                    Data block 1
     Block 2        Header block 2
                    Data block 2
     etc. for following blocks
```

File Header:

The File Header comprises 2 lines;

Line 1: the first 4 positions of the line must contain either the word FREE or the word FIXE:
FREE: the data will be read in free format (data separated by blank(s) or a comma)
FIXE: the data will be read in fixed format (each value has the same field length)

Line2: the first 3 positions of the line must contain either the word ROW,COL, or PAR:
ROW: one series per block and the data are time sequential row-wise
COL: one series per block and the data are time sequential column-wise
PAR: multiple series per block and the data are time sequential column-wise

Block Header(s)

The layout of the data Block Header depends beside on the layout of the data block also on the contents of the File Header; following distinction is made:

1. single series row-wise reading,see section 4.2.2.2
   1.a File header = FREE+ROW
   1.b File Header = FIXE+ROW,

2. single series column-wise reading,see section 4.2.2.3
   2.a File Header = FREE+COL
   2.b File Header = FIXE+COL

3. multiple series column-wise reading,see section 4.2.2.4
   3.a File Header = FREE+PAR
   3.b File Header = FIXE+PAR

4. row- or column-wise reading with data conversion by formula,see section 4.2.2.5

4.2.2.2 Row-wise data blocks

1. Block Header for data in free format (File Header is FREE+ROW):

```
First line :   Series code first series (fixed format):
     pos. 1 - 10: station code
         11 - 12: data type
              13:time interval unit
          14- 15:time interval divider
                  (right justified)

Second line :  (NOTE:this line must end with a slash!)
               -start date
               (date in HYMOS format:yyyy, mm, dd, hh, si)
               -number of rows
               -number of columns
               -conversion factor, if required
               (default = 1.)
               -reference level, if required
               (default= 0.)
               -/
```

NOTE

If the last record of the data block is incomplete (i.e. if the number of values is less than the number of columns), the record must end with a /(slash).

EXAMPLE : FREE+ROW

Layout of a data file consisting of two data blocks.

The first block comprises daily discharges of station ROORNIH01 from 1 January 1990 to 31 December 1990, i.e. 365 data. The number of columns is 10, hence the number of rows is 37 of which 36 are complete and the last one is incomplete.

The second block comprises monthly rainfall totals of station ROORNIH02 from April 1947 to March 1991, 12 values per row.

Data file	Comment
FREE	File Header line 1
ROW	File Header line 2
ROORNIH01 QH3 1	Header Blk.1 line1
1990,1,1,0,1,37,10,1.,0. /	Header Blk.1 line2
1.50,2.12,2.23,2.59,2.63,3.75,4.50,4.66,5.04,5.15	row 1
5.44,5.46,5.82,6.66,6.95,7.23,6.82,8.97,10.15,14.28	row 2
...	...
...	...
...	...
1.89,1.76,1.64,1.43,1.34 /	row 37
ROORNIH02 PH3 1	Header Blk.2 line1
19990,1,1,0,1,37,10,1.,0. /	Header Blk.2 line2
49 52 57 78 89 71 72 70 64 69 52 44	row 1
...	...
...	...
...	...
53 58 52 75 101 69 65 74 58 74 49 49	row 44

Note that the last row (37) of Block 1 is broken off with "/" because the row is incomplete (number of values is less than the number of columns), whereas in the last row (44) of Block 2 no "/" is applied as the number of data in the row corresponds with the number of columns.

2. Block Header for data in fixed format (File Header is FIXE-ROW):

First line: Series code (fixed format):

pos. 1	-	10:	Station code
11	-	12:	data type
		13:	time interval unit
14	-	15:	time interval divider (right justified)

Second line:
- start date (date in HYMOS format: yyyy,mm,dd,hh,si)
- number of rows
- number of columns
- conversion factor, if required (default = 1.)
- reference level, if required (default = 0.)
- start position of first data field
- length of data field (all fields must have the same length)

NOTE

1. In this FIXEd-format reading one must not apply a"/"to break off an incomplete row at the end of a data block! (This is in deviation from the FREE-format reading case shown above.)

2. Because the start position of the first data field in a row is to be specified, the left side of the data blocks before the start position of the first data field will not be read and may therefore be used for some codes on e.g. dates, see figure 5.

EXAMPLE FIXE+ROW

Layout of data file with hourly rainfall amounts of the station UORDOH. The data blocks for UORDOH contains 75 rows and 10 columns and starts on the first of May 1983; the data are given in mm which is the same as the unit of data type PH, hence the conversion factor is 1. and the reference level is 0.; the field length is 5 and the first field in a row starts at the first position.

Data file											Comment
FIXE											File Header line
ROW											File Header line
UORDOH PH4 1											Block Header Line
1983	5	1	0	1	75	10	1.	0.	1	5	Block Header line
	-1.0	-1.0	.4	.2	.3	.8	.0	.3	.0	.3	row 1
	.0	.0	.0	.0	.0	.0	3.2	.0	.0	.1	row 2
	.1	.0	.0	.0	.0	.0	.0	.0	.0	.0	row 3
	.0	.0	.0	1.3	1.5	.0	.0	.0	.0	.0	row 4
	.7	.0	1.2	.3	.0	.0	.0	.0	.0	.0	row 5
	.0	.0	.2	.2	.5	.2	.0	.0	.0	.5	row 6
...											...
...											...
...											...
	.0	.0	.0	.0	.0	.0	3.2	.0	.0	.1	row 74
	.0	.0	.0	.0							row 75

4.2.2.3 Column-wise data blocks

1. Block Header for data in free format (file Header is FREE+COL):

First line: Series code (fixed format):

pos. 1 - 10: station code
11 - 12: data type
13: time interval unit
14 - 15: time interval divider (right justified)

Second line: (NOTE: this line must end which a slash!)

- start date (date in HYMOS format: yyyy,mm,dd,hh,si)
- maximum number of rows (nrows)
- number of columns (ncolumns)
- conversion factor, if required (default = 1.)
- reference level, if required (default = 0.)
- /

Third line: number of rows in each column

Next lines: data blocks of size (nrows*ncolumns).

NOTE

The columns should be completed with some value (e.g. 0) up to the maximum number of rows in a column, (see also columns 2 and 5 in the example shown below).

EXAMPLE FREE+COL

Layout of data file of daily discharges of station ROORIRI from 1 October 1980 to 31 March 1981. The maximum number of rows in a column is 31 (max. number of days in a month) and the number of columns is 6 = number of months. Note, as the column 2 (= November) and 5 (= February) are incomplete their columns are filled up with zero's for reading purposes; these zero's are not transferred to the database.

Data file	Comment
FREE	File Header line
COL	File Header line
ROORIRI QQ3 1	Block Header line
1980 10 1 0 1 31 6 1 0 /	Block Header line
31 30 31 31 28 31	Block Header line
.10 .34 .52 .54 .39 .31	row 1
.10 .29 .49 1.11 .38 .41	row 2
...	...
...	...
...	...
.21 .25 .45 .58 0 .46	row 30
.34 0 .38 .52 0 .60	row 31

2. Block Header for data in fixed format (File Header is FIXE-COL):

First line: Series code (fixed format):
- pos. 1 - 10: Station code
- 11 - 12: data type
- 13: time interval unit
- 14 - 15: time interval divider (right justified)

Second line: (NOTE: this line must end which a slash!)
- start date
  (date in HYMOS format: yyyy,mm,dd,hh,si)

- maximum number of rows (nrows)
- number of columns (ncolumns)
- conversion factor, if required (default = 1.)
- reference level, if required (default = 0.)
- start position of first data field
- length of data field

Third line: number of rows in each column

Next lines: data blocks of size (nrows*ncolumns).

EXAMPLE FIXE+COL

Layout of data file of daily discharges of station ROORIRI from 1 October 1980 to 31 March 1981, almost similar to the free format case shown above. The maximum number of rows in a column is 31 (max. number of days in a month) and the number of columns is 6 = number of months. The first two positions in each line are used for comment (day number) and need not be read, hence the first data field starts at position 3. The field length is 8.Note that as the reading is in FIXEd format the columns need not be completed with zero's as was the case with FREE format reading in the example above.

Data file	Comment
FIXE	File Header line 1
COL	File Header line 2
ROORIRI QQ3 1	Block Header line1
1980 10 1 0 1 31 6 1 0 3 5	Block Header line2
31 30 31 31 28 31	Block Header line3
1 .10 .34 .52 .54 .39 .31	row 1
2 .10 .29 .49 1.11 .38 .41	row 2
3 .10 .26 .47 2.64 1.05 .50	row 3
...	...
...	...
...	...
30 .21 .25 .45 .58 .46	row 30
31 .34 .38 .52 .60	row 31

4.2.2.4 Multiple series data blocks

If data blocks comprise more than one series, each column representing a different series, HYMOS can read the block:

-format free: then all column will be transferred, i.e. number of columns = number of series, or

-format fixed:then not all columns are necessarily transferred to the database, i.e. number of columns $\geq$ number or series.

This leads to the following Block Headers.

1. Block Header for data in free format (File Header is FREE+PAR):

First line: Number of series N_s in the data block (=number of columns)

Next N_s-lines: Series code(s),(fixed format),one per line:

pos. 1 - 10 : station code
11 - 12 : data type
13 : time interval unit
14 - 15 : time interval divider (right justified)

Next line: (NOTE:this line must end with a slash!)
- -start date (date in HYMOS format: yyyy, mm, dd, hh, si)
- -number of rows (nrows),of equal length for all columns in the block
- -number of columns(ncolumns),must be equal to N_s, the number of series
- -conversion factor(if required), the same for all series (default = 1.)
- -reference level (if required), the same for all series (default = 0.)
- - /

Next lines: data block of size (nrows*ncolumns).

EXAMPLE FREE+PAR

Layout of data file of daily discharges of stations RISHI HARIDO1 HARIDO2 AND ROORIRI from 1 October to 30 November 1980. Hence there are 4 columns with length 61 (total number of days in two months).

Data file	Comment
FREE	File Header line 1
PAR	File Header line 2
4	Block Header line1
RISHI QQ3 1	Block Header line2
HARID01 QQ3 1	Block Header line3
HARID02 QQ3 1	Block Header line4
ROORIRI QQ3 1	Block Header line5
1980 10 1 0 1 61 4 1. 0. /	Block Header line6
.10 .34 .52 .54	row 1
.10 .29 .49 1.11	row 2
.10 .26 .47 2.64	row 3
.10 .24 .46 1.31	row 4
...	...
...	...
...	...
.21 .25 .45 .58	row 60
.34 .29 .38 .52	row 61

2. Block Header for data in fixed format (File Header is FIXE+PAR):

First line: Number of series N_s in the data block to be transferred (≤ number of columns)

Next N_s-lines: Series code(s), start position and field length (fixed format), one per line:

pos. 1 - 10 : station code
11 - 12 : data type
13 : time interval unit
14 - 15 : time interval divider (right justified)
16 - 17: start position of series field (right justified)
18 - 19 : field length (right justified)

Next line:
-start date (date in HYMOS format: yyyy, mm, dd, hh, si)
-number of rows (nrows), of equal length for all columns in the block
-conversion factor (if required), the same for all series (default = 1.)

-reference level (if required), the same for all series (default = 0.)

Next lines: data block of size (nrows*ncolumns).

EXAMPLE FIXE+PAR

Layout of data file of daily discharges of stations RISHI, HARID01, HARID02 and ROORIRI from 1 October to 30 November 1980. The series of station HARID01 and ROORIRI will be transferred. In the file the first 2 positions of each line are reserved for comments. The field lengths of the series are respectively:

-discharge RISHI, field length = 7,
-discharge HARID01, field length = 5,
-discharge HARID02, field length = 5,
-discharge ROORIRI, field length = 5,

Hence the series of station HARID01 start at position 2+7+1=11 and the one of station ROORIRI at position 2+7+5+5+1=20. The column length is 61 (total number of days in the two months).

Data file	Comment
FIXE	File Header line 1
PAR	File Header line 2
2	Block Header line1
HARID01 QQ3 110 5	Block Header line2
ROORIRI QQ3 120 5	Block Header line3
1980 10 1 0 1 61 1. 0.	Block Header line4
01 .10 .34 .52 .54	row 1
02 .10 .29 .49 1.11	row 2
03 .10 .26 .47 2.64	row 3
04 .10 .24 .46 1.31	row 4
...	...
Table contd. ...	
...	...
...	...
60 .21 .25 .45 .58	row 60
61 .34 .29 .38 .52	row 61

4.2.2.5 Transfers with polynomial conversions

A polynomial type of conversion of data in a block, while transferring data to the HYMOS database, can only be applied to data blocks comprising one series, i.e. the cases where the second line in the file header reads ROW or COL. If a polynomials conversion is requested, then the Block Header is as follows:

First line: Series code (fixed format):

pos. 1	-	10:	Station code
11	-	12:	data type
		13:	time interval unit
14	-	15:	time interval divider (right justified)
16	-	22:	the word FORMULA

Second line: as for row-or column-wise time sequential reading with dummies for conversion factor and reference level

Third line: coefficients c_o, c_1, c_2, c_3 and c_4 (all of them and separated by commas) of the 4th order polynomial:

$$Y = c_o + c_1X + c_2X^2 + c_3X^3 + c_4X^4$$

The remainder is as before.

EXAMPLE

A data logger at station ROORIRI stores gate levels with intervals of 20 minutes. To transform the logger units to SI-units in meters above gauge zero a polynomial of the form is used:

$$H_{gate} = 3.675 + 4.354X + 0.3098X^2 - 3.505X^3 + 2.3966X^4$$

The logger units are in fixed format row-wise time sequential, 3285 rows and 8 columns. The first field in a row starts at the first position and the field length is 6. The first value is at 00.00 hrs 1 January 1990.

Data file	Comment
FIXE	File Header line 1
ROW	File Header line 2
ROORIRI HK4 3FORMULA	Block Header line1
1990,1,1,0,1,3285,8,1,0,1,8	Block Header line2
3.675 4.3354 0.3098 -3.505 2.3966	Block Header line3
.3884 .3884 .3884 .3884 .3884 .3884 .3884 .3884	row 1
.3884 .3884 .3884 .3884 .3884 .3884 .3884 .3884	row 2
.3884 .3884 .3884 .3884 .3884 .3884 .3884 .3884	row 3
.3884 .3884 .3884 .3884 .3884 .3884 .3884 .3884	row 4
etc.	

4.2.3 Data files without headers

Data files without headers requires data block(s):

1. in fixed format, and
2. with the data time sequential row-wise.

The series code, start date, block layout, conversion factor, reference level, start position of first data field and length of the data field are then entered from the screen.

4.3 Non-equidistant time series

The data files of non-equidistant time series comprise only one data block, which may contain a number of series, structured column-wise. For entry to the HYMOS data base a non-equidistant time series data file should have the following layout:

Line 1 to 10

pos. 1 - 10: station code of first series

11 - 12: data type of first series followed by similar declarations for the other data series on lines 2 to 10.

In case it is required to skip certain columns or the file contains less than 10 data columns, leave the corresponding lines open. A maximum of 10 columns of data can be loaded from the same file.

Next lines: per line (free format):
- year,
- month,
- day,
- hour,
- minute,
- value series 1,
- value series 2,
  etc.
- /

NOTE

It is not allowed to enter a column which contains empty spaces. In that case, the empty spaces have to be filled with for example the selected value for missing data!

EXAMPLE

Given is a water quality parameter data file of station ROORIRI with the following layout:

date	O_2	KjN	NH4	NO2	NO3	C1
80-01-07	7.75	4.3	2.6	0.09	4.8	72
80-01-22	10.50	6.2	5.5	0.08	2.9	108
80-02-06	7.20	8.1	3.6	0.09	5.3	65
80-02-21	9.65	6.6	3.7	0.11	4.2	90
80-03-07	9.60	5.7	5.2	0.15	3.2	91
80-03-17	8.80	5.1	2.7	0.07	4.0	73
etc.						

The series codes are respectively for:

02	=	ROORIRI 02
KjN	=	ROORIRI KN
NH4	=	ROORIRI N4
NO2	=	ROORIRI N2
NO3	=	ROORIRI N3
CL	=	ROORIRI CL

To transfer these series to the HYMOS database the file should have the following layout:

ROORIRI   02

```
ROORIRI    KN
ROORIRI    N4
ROORIRI    N2
ROORIRI    N3
ROORIRI    CL
80 01 07 0 0    7.75    4.3    2.6    0.09    4.8     72 /
80 01 22 0 0   10.50    6.2    5.5    0.08    2.9    108 /
80 02 06 0 0    7.20    8.1    3.6    0.09    5.3     65 /
80 02 21 0 0    9.65    6.6    3.7    0.11    4.2     90 /
80 03 07 0 0    9.60    5.7    5.2    0.15    3.2     91 /
80 03 17 0 0    8.80    5.1    2.7    0.07    4.0     73 /
etc.
```

4.4 Stage-discharge data

The stage-discharge data file should have following layout:

First line: Station code (format A10)

Next line:
- year, month, day of the measurement
- number of the measurement
- gauge zero (m + M.S.L.)
- water level (m)
- discharge (m^3/s)
- gradient (m/day) or fall (m)
- width (m)
- wetted perimeter (m)
- cross-sectional area (m^2)

Etc. for the next lines.

NOTE

1. All data should be separated by comma's or blanks. To break off a record (if e.g. the last items are not available) use "/"; then the omitted values will be set to the default = 0.
2. The data should be in SI-units.

EXAMPLE

Stage-discharge data of station ROORIRI including fall. The records are broken off after the fall. The water levels in this file as well as in the time series are given relative to gauge zero; hence no shift is required and the gauge zero can remain 0.0. The layout of the data file is shown below.

```
ROORIRI
1990  1  1  1 0.0    5.907  1160  1.917/
1990  1  3  2 0.0    7.105  1520  2.182/
1990  1  3  3 0.0    5.026   889  1.597/
1990  1  8  4 0.0    7.013  1490  2.225/
1990  1 10  5 0.0   11.558  2830  2.880/
1990  1 11  6 0.0    8.108  1640  1.920/
1990  1 17  7 0.0    8.638  1990  2.652/
1990  1 18  8 0.0    3.139   399   .808/
etc.
```

5.0 Data Validation

5.1 Screening

5.1.1 General

As a first step in data validation screening of data is performed to obtain proper listings of series for easy reference and first checks on the range of data. HYMOS provides following options for first validation of series:

1. listing of series
2. screening of series,
3. comparison of series, and
4. tabulation of series.

The first option applies to equidistant and non-equidistant series; all other options can only be executed for equidistant time series.

The options are elaborated in the following sections.

5.1.2 Listing of series

Series can quickly be presented in dedicated tables and data blocks.

Dedicated tables:

For equidistant time series with the following time intervals tables with a dedicated layout can be obtained:

- 15 minutes,
- hour,
- day,
- decade

- month, and
- annual

Data blocks:

For equidistant and non-equidistant time series the data can be presented in data blocks by prescribing the number of rows and columns on one page.

Further options include:

- specification of number of decimals in the tables,
- coding of data:
  - ' ' = original data
  - '*' = series' element is completed
  - '+' = series'element is corrected

Examples of listings are presented in Tables 1 and 2.

5.1.3 Screening of series

Screening of series comprises listing of equidistant time series, combined with:

1. computation of statistics, and
2. marking of outliers.

To what extent the options apply depends on the choice of the table.

In case of dedicated tables this includes:

- statistics:
  - effective number of data
  - number of data missing
  - sum and mean
  - standard deviation
  - coefficient of variation
  - minimum
  - maximum
  - number of data below lower boundary
  - number of data above upper boundary
- marking of outliers: data exceeding boundaries will be marked with '+'. The boundaries are determined in one of the following ways:
  - based on given boundaries: X^+, X^-
  - relative boundaries: $X^+ = m_X + \alpha . s_X$

    $X^- = m_X - \beta . s_X$

    where α and β are input and m_X and s_X refer to mean and standard deviation of series X; e.g. for monthly series these quantities refer to monthly mean and standard deviation.

In case of data blocks no statistics will be computed; only outliers exceeding the given boundaries X^+ and X^- will be marked.

An example of the screening option is presented in Table 3.

5.1.4 Comparison of series

The comparison option includes the selection of pairs of series with the same time interval. It provides the possibility to obtain a quick summary of differences between the two series. A table is provided showing the data for the time-steps the series differed. This option is particularly useful when differences between updated/corrected series and their original have to be detected.

5.1.5 Tabulation of series

A listing is presented of up to 6 time series side by side. This technique is particularly useful to trace shifts between series , e.g. for rainfall data. These shifts may e.g. be caused by errors in dating of the data. Well known is the day shift for daily precipitation.

In Table 4 an example of this option for daily rainfall data is presented.

5.2 Time series graphs

5.2.1 General

The <time series graphs> option include graphs of:

- time series,
- residual series,
- residual mass curves,
- moving averages,
- water balances,
- barcharts of series availability.

These main options are typically meant for data validation purposes. Graphics for reports can better be made via the <graphics> option under <reporting>. Nevertheless, <time series graphs> provides a number of user options to improve the layout of graph; these options are listed in section 5.2.2. The main options are dealt with in the sections 5.2.3 to 5.2.8.

5.2.2 Graphical options

The flexibility in the creation of graphs is enhanced by following options:

- plot series: plotting of basic series together with residual mass curve, moving averages or water balance;
- size: A3 or A4 size;
- scale: setting of minimum, maximum and step size of vertical axis; in case series are of different type two axis can be defined;
- shift: series can be shifted along the vertical in units of the series, whereas one or more series may also be scaled by means of a multiplier (default = 1.0);
- device: plot device: - screen,
  - plotter, and
  - file;
- cadre text: title, sub-title, company name, figure number, etc. may be added to the graph;
- function text: text for time and function axis as well as for the series can be added (defaults are available);
- graph type: series may be presented by lines or bars,(lines is default);
- line type: lines can be coloured, coded or all equal. In Figure 6 some of the options are illustrated.After a plot is made one can select one of the following options:

  - main menu: back the first screen from where the main option can be selected;
  - series: for the same main option other series can be selected;
  - period: the time period of the selected series, for which a plot is required, can be altered.
  - device: after having seen the plot on the screen one can decide to send it to the plotter or to a file.
  - option: this refers to a change in one of the plot options discussed above for the

selected series and time period.

5.2.3 Time series

Up to 5 series can be plotted in one graph. All options presented in section 5.2.1 apply. Some examples of graphs, showing the flexibility of the option are shown in figures 7 and 8.

5.2.4 Residual series

A residual series is a series plotted relative to the mean value of the series. The residual series gives a quick insight in wet and dry periods. This shows best if the series is plotted as bars. Although not strictly necessary best is to plot only one series per graph. An example of residual series is shown in Figure 9.

5.2.5 Residual mass curve

A residual mass curve represents accumulative departures from the mean. It is an efficient tool e.g. to detect climatic variabilities or other inhomogeneities. The residual mass curve Y_i is derived as follows:

$$Y_{x,i} = Y_{x,i-1} + (X_i - m_x) = \sum_{j=1}^{i} (X_j - \frac{1}{N} \sum_{k=1}^{N} X_k) \quad \ldots \quad (1)$$

where: N ⇒ number of elements in the series

The curve can be interpreted as follows:

- an upward curve indicates an above average sequence
- a horizontal curve indicates an about average sequence, and
- a downward curve indicates a below average period

The original series X_i can be plotted together with the residual mass curve if so required, by selection of <plot series>.

An example is shown in Figure 10.

5.2.6 Moving Average

To investigate the long term variability or trends in series moving average curves are useful. A moving average series Y_i of series X_i is derived as follows:

$$Y_{x,i} = \frac{1}{2M + 1} \sum_{j=i-M}^{i+M} X_j \qquad (2)$$

where averaging takes place over 2M+1 elements. An element of the moving average series gets a missing value if one of the X_j is missing.

The original series X_i can be plotted together with the moving average series if so required, by selection of :<plot series>.

An example of a moving average series is presented in Figure 11.

5.2.7 Balance

To check the consistency of series the mass conservation condition is often a useful investigative tool. This applies in particular to discharge series. Basically, water balances are made of discharge series of successive stations along a river or of stations around a junction, where there should be a surplus, a balance or a deficit, depending on water is added or lost. To give full flexibility in using this option signs and multipliers to series values are input. At maximum 4 series can be selected, The balance equation $Y_i = F(X_{j,i})$ reads:

$$Y_i = \pm\, a.X_{1,i} \pm b.X_{2,i} \pm c.X_{3,i} \pm d.X_{4,i} \qquad \ldots(3)$$

where : a,b,c,d = multipliers entered by user (default = 1)
$\pm$ = sign entered by user (default = +)

In the balance plot two function axes, one for the X_j's and one for Y_i can be specified. An example is given in Figure 12.

5.2.8 Series availability

The availability of equidistant and non-equidistant time series can be presented in a bar chart. The availability of data of up to 20 series can be shown in one chart. In case of equidistant time series all series should have the same time interval. The series can be selected randomly, from the displayed scrolling menu, or sequentially. In the latter case the sequence corresponds with the sequence in the scrolling menu; the first and last series code have to be indicated.

In the legend of the barchart symbols for missing data and empty vectors (space where no data are loaded yet) are given.

In case of non-equidistant time series the number of data per calendar unit are presented.

An example of a barchart of series availability is presented in Figure 13.

5.3 Relation curves

5.3.1 General

A relation curve gives a functional relationship between two series of the following form $Y_t=F(X_{t+t1})$. The curves can be used for :

1. detection of random errors,
2. detection of systematic errors,
3. filling in of missing data, and
4. forecasting purposes,

If there is a strong one to one relationship between two series random errors will be shown in a relation curve plot as outliers. To arrive at a one to one relationship (i.e;elimination of loopings) the introduction of a time shift (t1) between the two series may be necessary.

By comparing two relation curves or data of one period with the curve of another period, shifts in relationships, e.g. in water level series due to changes in the gauge zero, can be detected.

The relation curve fitted to the data of two series can be used to fill-in missing data in the dependent variable of the relation (Y) (also see section 6.1 on Interpolation).

If the series in the relation are mutually shifted in time, with sufficient lead-time for the independent variable X(t1 negative), the relation curve may be used to forecast the dependent variable in the relation Y from observations on X.

The parameters of the established relationships for a period of time can be stored in the data base for e.g. later comparison, filling-in missing data.

The main options under <relation curves>includes:

- optimization of time shift t1,
- plotting of time series data Y_t versus X_{t+t1},
- fitting a polynomial to Y_t, X_{t+t1},
- validation of relation curve,
- display and comparison of relation curves.

5.4 Double mass analysis

5.4.1 General

Double mass analysis is a technique to detect possible inhomogeneities in series, like jumps, trends, etc. by investigating the ratio of accumulated values of two series, viz:

- the series to be tested, and
- the base series.

The base series is generally an artificial series, i.e. the average of reliable series of nearby stations, which is assumed to be homogeneous.

The result of the analysis is presented in a table and in a plot: the double mass curve. This curve will show a straight line if the test-series is homogeneous. A jump in the test-series will create a break in the double mass curve, whereas a trend will create a curved line.

The computational is procedure presented in section 5.4.2.

5.4.2 Computational procedure

Let X_i, (i = 1,N) be the test series and Y_i,(i=1,N) the base series. The double mass analysis then considers the following ratio:

$$rc_i = \frac{\sum_{j=1}^{i} X_j}{\sum_{j=1}^{i} Y_j} \qquad (4)$$

or expressed as a ratio of the percentages of the totals for N elements :

$$pc_i = \frac{\sum_{j=1}^{i} X_j}{\sum_{j=1}^{N} X_j} \; \frac{\sum_{j=1}^{N} Y_j}{\sum_{j=1}^{i} Y_j} \qquad (5)$$

In the tabular results 9 columns are presented:

1. time
2. value of series X
3. accumulated value of series X
4. accumulated value as a percentage of the total of X
5. value of series Y
6. accumulated values of series Y
7. accumulated value as a percentage of the total of Y
8. ratio (item 3)/(item 6), equation 4
9. ratio (item 4)/(item 7), equation 5

The last column is also presented in the double mass plot. Examples of table and plot are shown in Table 5 and Figure 14 respectively.

If the curve shows a distinct break with curve slopes α before and β after the break, adjustment may take place in two ways:

- either the data before the break are adjusted to the present conditions by multiplication by the ratio : β/α, or
- the data after the break are adjusted to the pre-break conditions : in that case the recent data are multiplied by a factor α/β.

The correction can be materialised by the transformation option of HYMOS under <data complication >.

5.5 Series homogeneity tests

5.5.1 General

Dependent on the type of analysis series must fulfill one or more of the following requirements :

- stationarity: i.e. the properties or characteristics of the series do not vary with time,
- homogeneity: i.e. all elements of a series belong to the same population ;
- randomness: i.e. series elements are independent.

HYMOS includes following statistical tests to investigate series' stationarity, homogeneity or randomness :

1. Median run test: a test for randomness by calculating the number of runs above and below the median ;
2. Turning point test : a test for randomness by calculating the number of turning points;
3. Difference sign test :a test for randomness by calculating the number of positive and negative differences;
4. Spearman rank correlation test :the Spearman rank correlation coefficient is computed to test,
   -the existence of correlation between two series,
   -the significance of serial rank correlation, and
   -the significance of a trend;
5. Arithmetic serial correlation coefficient: a test for serial correlation;
6. Wilcoxon-Mann-Whitney U-test: a test to investigate whether two series are from the same population;
7. Student t-test : a test on difference in the mean between two series ;
8. Wilcoxon W-test : a test on difference in the mean between two series ;
9. Linear trend test :a test on significance of linear trend by statistical inference on slope of trend line ;
10. Range test :a test for series homogeneity by the rescaled adjusted range.

Notes:

1. The Spearman rank correlation test may be used as a single or two series test; in the single series mode it tests the significance of correlation with time.
2. Tests nrs. 6,7 and 8 (Wilcoxon-Mann-Whitney U-test, Student t-test and Wilcoxon W-test) are basically two series tests;

however, the test can also be used for a single series by means of the split-sample approach, where a series is divided into two parts, which are mutually compared.

5.6 Spatial homogeneity test

5.6.1 General

The test described here is applicable to quality and quantity parameters with a spatial character, like rainfall, temperature, evaporation, etc., but sampled at a number of stations (point measurements).

To investigate the reliability of point observations at a station, called the base station, the observations are compared with weighted averages of the rainfall at neighbor stations. The weights are inversely proportional to some power of the distance between the base station and the neighboring stations. The test considers the difference between the observed and estimated values at the base station. If the absolute difference between observation and estimate exceeds specific limits (absolute and relative), the observation will be flagged out (not deleted) to stress the need for further investigation.

5.6.2 Test procedure

To be specific and to avoid general phrasing the test will be explained for rainfall series, where rainfall may be replaced by any other spatial parameter.

In this section the following topics will be discussed:

- selection of neighbor stations
- estimation of point rainfall
- test criterion
- corrections for heterogeneity
- limitations

Selection of neighbor stations

Following criteria are used to select the neighbors of the base station:

1. series with the same data type and interval as the one under investigation should be available
2. the distance between the basic station and a neighbour should be less than a specified maximum correlation distance

R_{max} (km);

3. maximum amount of neighbors is 8;
4. per quadrant at maximum 4 stations out of 4 fulfilling criterion 2 can be selected , see Figure 15, but criterion 3 remains valid, default are the two stations nearest to the base station.

The selection on quadrants is applied to obtain a proper spatial distribution of stations around the base station. However, due to prevailing wind conditions or orographical effects spatial heterogeneity may be present. In those cases normalized rather than actual values should be investigated.

Estimation of point rainfall:

The point estimate for the base station based on the observations at N neighbour stations for the same time interval reads :

$$P_{est}(t) = \frac{\sum_{i=1}^{N} P_i(t)/D_i^b}{\sum_{i=1}^{N} 1/D_i^b} \qquad (6)$$

where:

$P_{est}(t)$ = estimated rainfall at the base station at time t

$P_i(t)$ = measured rainfall at neighbour station i at time t

D_i = distance between the base station and neighbour station i

N = number of neighbour stations taken into account

b = power of distance D, (usually b = 2).

Test criterion:

The difference between the observed value, $P_{meas}(t)$, and the estimated value, $P_{est}(t)$, is considered to be insignificant if the following conditions are met :

$$| P_{meas}(t) - P_{est}(t) | \leq X_{abs} \quad \text{, and} \qquad (7)$$

$$\mid P_{meas}(t) - P_{est}(t) \mid \leq X_{rel} * S_{P_{est}(t)} \quad (8)$$

with :

X_{abs} =admissible absolute difference

$S_{pest(t)}$ = standard deviation of neighbouring values, see equation (9)

X_{rel} = multiplier of standard deviation

$$S^2_{P_{est}}(t) = \frac{1}{N} \sum_{i=1}^{N} (P_i(t) - \overline{P_i(t)})^2 \quad (9)$$

If the difference is unacceptably high, the recorded value is flagged "+" or "-", depending on whether the observed total is greater or less than the estimated. In case no estimate is available the value will be flagged with "*". Only the flagged , i.e. the suspicious, data will be printed.

Corrections for heterogeneity:

To correct for sources of heterogeneity, e.g. orographical effects. normalised rather then actual values may be used. This implies that in the equations (6) through (9) the observations at the neighbour stations are multiplied by the ratio of the base station normal and the neighbour station normal :

$$P_{ci} = (N_{base}/N_i).P_i \quad (10)$$

where :

P_{ci} =for heterogeneity corrected value at neighbour station i

N_{base} =normal of base station

N_i =normal of neighbour station i

The station normals are read from a station-normal file or are given as a function of the altitude of a station.

The data for the former option can be entered via <station normals> of <input data files> from the <data entry & editing> menu.

The station normal as a function of the station altitude are of the following form :

$$N_i = a_1+b_1.H_s \qquad \text{for } H_s<H_1 \qquad (11)$$

$$N_i = a_2+b_2.H_s \qquad \text{for } H_s>H_1 \qquad (12)$$

Limitations:

Following limitations to the use of the spatial validation option apply:

1. Only series of the same type are considered
2. Only series with the same time interval are considered
3. Generally, all series available in the data base of the selected data type and time interval will be considered. However, if station normals are read from a station-normal file, then only those stations will be considered that are available in that file.

6.0 Series completion and regression

6.1 Interpolation

6.1.1 General

To fill - in missing data HYMOS offers following interpolation methods :

- linear interpolation
- use of relation equations
- spatial interpolation

The methods are presented in the sections 6.1.2 to 6.1.4

6.1.2 Linear interpolation

In a number of cases gaps in series can well be filled in by linear interpolation between the last value before the gap and the first one after, provided that the distance over which the interpolation takes place is not too large.

The use of this option requires following input :

- selection of series interval,
- selection of series (from a scrolling menu),
- the period to be considered for filling -in , and
- the maximum interpolation distance (expressed as a number of time intervals). This means that gaps larger than this maximum will not be filled -in.

6.1.3 Relation equation

Relation/regression equations can be used to fill-in missing data, provided that the standard error in the fit is small. Since relations between stations may change seasonally, HYMOS offers the option to apply a relation only to a period in the year, to get the best estimate possible for the data gap.

The following types of equations may be used to fill-in missing data :

- -Polynominal
- -Simple linear
- -Expotential
- -Power
- -Logarithimic
- -Hyperbolic
- -Multiple linear

6.1.4 Spatial interpolation

The spatial interpolation technique is applicable to quality and quantity parameters with a spatial character, like rainfall, temperature, evaporation, etc., but sampled at a number of stations (point measurements). Missing data at a base station are estimated by weighted averages on observations at neighbour stations. The weights are inversely proportional with some power of the distance between the base station and the neighbour stations.

Selection of neighbour stations:

Following criteria are used to select the neighbours of the base stations :

1. series with the same data type and interval as the base station series should be available;
2. the distance between the basic station and a neighbour should be less than a specified maximum correlation distance R_{max} (km)
3. maximum amount of neighbours is 8;
4. per quadrant at maximum 4 stations out of 4 fulfilling criterion 2 can be selected, see Figure 15, but criterion 3 remains valid. Default are the two stations nearest to the base station.

The selection on quadrants is applied to obtain a proper spatial distribution of stations around the base station.

However, e.g. for rainfall, due to prevailing wind conditions or orographical effects spatial heterogeneity may be present. In those cases normalized rather than actual values should be used.

Estimation of point rainfall:

The point estimate for the base station based on the observations at N neighbour stations for the same time interval is given by equation 6.

Corrections for heterogeneity:

To correct for sources of heterogeneity, e.g. orographical effects, normalised rather then actual values may be used. This implies that in the equation (6) the observations at the neighbour stations are multiplied by the ratio of the base station normal and the neighbour station normal, see equation (10). The station normals are read from a station-normal file or are given as a function of the altitude of a station. The data for the former option can be entered via <station normals> or <input data files> from the <data entry & editing> menu.

The station normal as a function of the station altitude are given by equation 12.

6.2 Regression analysis

6.2.1 General

The regression analysis option in HYMOS includes:
- computation of correlation matrix, and
- fitting of following type of functions
  - polynomial
  - simple linear
  - exponential
  - power
  - logarithmic
  - hyperbolic, and
  - multiple linear.

The multiple linear functions can be fitted by means of multiple or stepwise regression techniques.

The main steps in running the regression option include:
- input: to enter the data set,
- function: to select one of the available functions,

-select: to select the series in regression, and
-store: to store parameters in the data base.

6.2.2 Regression equations

The following types of regression equations are available, with Y the dependent variable and X_j's the independent variables:

1. polynomial

$$Y_i = \sum_{j=0}^{n} C_j X_i^n \qquad (13)$$

with : n = degree of polynomial : $n \leq 9$,

C_j = coefficient

2. simple linear

$$Y_i = A + B.X_i \qquad (14)$$

with : A,B = coefficients

3. exponential

$$Y_i = A \exp (B.X_i) \qquad (15)$$

with: A,B - coefficients

4. exponential

$$Y_i = A \exp (B/X_i) \qquad (16)$$

with : A,B - coefficients

5. power

$$Y_i = A.X_i^B \qquad (17)$$

with: A,B = coefficients

6. logarithmic

$$Y_i = A + B.\ln(X_i) \qquad (18)$$

with: A, B = coefficients

6.3 Rainfall-runoff simulation

7.3.1 General

The simulation of the rainfall -runoff process in a catchment aims at

- filling-in and extension of discharge series,
- generation of discharges from synthetic rainfall,
- real time forecasting of flood waves, and
- determination of the influence of a changing land/water use, or more generally, determination of the influence of human interference in the runoff process like the effect of subtraction and supply of water from or to the groundwater reservoir and the analysis of the influence of hydraulic structures on the runoff process (dams, dikes, reservoirs), in the design as well as in the management phase of water resources systems.

The latter aim can only be realized if the simulation model possess a distinct physical basis. The rainfall-runoff process in a catchment is extremely complex. A very detailed model, however, requires large amounts of information and computer time. Detailed information is usually not available, and a schematization of the rainfall- runoff process therefore is usually inevitable.

The Sacramento model offers a good compromise between (1) the physical background required (2), the amount of information available and (3) computational speed to simulate the runoff process in a catchment for a large number of years. To some extent also effects of human interferences can be incorporated, at least qualitatively.

In the simulation of the runoff process by the Sacramento model a distinction is made between the land phase and the channel phase.

The landphase is represented by an explicit moisture accounting lumped parameter model. The catchment area is divided into one or more segments discharging to the main channels, see Figure 16.

Within every segment areal homogeneity with respect to rainfall and basin characteristics is assumed. The distance is not explicitly treated in the simulation process as is time. In this respect the lumped approach deviates from the distributed catchment models.

The propagation and attenuation of floodwaves in the channel may be simulated by methods with varying degree of sophistication, ranging from simple summation via unit hydrograph methods to layered Muskingum approaches. The available methods are not very accurate for the simulation of the propagation of flood waves through the channel system. It is recommended to apply a dynamic hydraulic model for characteristic parts of the hydrograph, which are of importance for design purposes. The outflow from the landphase of the Sacramento model can then act as upstream and lateral inflow to the physically based channel model.

7.1 Processing of flow measurements

7.1.1 General

HYMOS can treat two types of flow measuring methods:

1. point velocity method
2. moving boat method

In both cases average velocities in a number of verticals in the cross-section are computed. From these velocities the discharge is computed by:

1. mean-section method
2. mid-section method

The computational results are presented in tables 6 and Figure 17. The condensed results are stored in the database.

7.2 Fitting of Rating Curve

7.2.1 General

HYMOS includes following options to fit stage-discharge data by a rating curve:

1. simple rating curve
2. rating curve with unsteady flow correction
3. rating curve with backwater correction.

The rating curve parameters are stored in the database. A rating curve is valid for a certain period of time. Each curve can be described by at maximum 3 sets of parameters valid for a specific water level-range. The curves may be of the parabolic or of the power type equation.

Full reports of the quality of fit and linear and double-logarithmic scale plots of the stage-discharge data and rating curve can be obtained.

In addition to the above options a procedure is included to qualify shift adjustments on water levels to account for river bottom variations in the stage-discharge conversion.

7.2.2 Before computation

A print-out of the stage-discharge data together with linear and double logarithmic plots are useful prior to the determination of the parameters of the rating curve:

- to check the availability of data within water-level ranges, and
- to investigate distinct breaks in the double logarithmic stage-discharge plot, which marks the range of applicability of sets of parameters in the rating curve.

It is advised to use approximately the same amount of data points per unit of depth in the determination of a set of parameters of a rating equation applicable to a certain water level range.

7.2.3 Simple rating curve

When unsteady flow and backwater effects are negligibly small the stage-discharge data can be fitted by a simple relationship, valid for a given period of time and water level range.

Following rating equations can be applied:

- parabolic type:

  for $h_i < h \leq h_{i+1}$: $\quad Q = a_1 + b_1 h + c_1 h^2$ (19)

- power type:

  for $h_1 < h \leq h_{i+1}$: $\quad Q = c_2 (h + a_2)^{b_2}$ (20)

where:

Q = discharge (m^3/s)

a_i, b_i, c_i = parameters

h_i, h_{i+1} = lower and upper water level for which the rating equation applies.

The coefficients a_1, b_1 and c_1 of the parabolic equation are determined by the least squares method. The shift parameter a_2 in the power equation is either input or determined by an adapted Johnson method. A computerized Johnson method (see for a description of the method e.g. WMO Operational Hydrology Report bottom variations in the stage-discharge conversion.

7.2.2 Before computation

A print-out of the stage-discharge data together with linear and double logarithmic plots are useful prior to the determination of the parameters of the rating curve:

- to check the availability of data within water-level ranges, and
- to investigate distinct breaks in the double logarithmic stage-discharge plot, which marks the range of applicability of sets of parameters in the rating curve.

It is advised to use approximately the same amount of data points per unit of depth in the determination of a set of parameters of a rating equation applicable to a certain water level range.

7.2.3 Simple rating curve

When unsteady flow and backwater effects are negligibly small the stage-discharge data can be fitted by a simple relationship, valid for a given period of time and water level range.

Following rating equations can be applied:

- parabolic type:

$$\text{for } h_i < h \leq h_{i+1}: \quad Q = a_1 + b_1 h + c_1 h^2 \tag{19}$$

- power type:

$$\text{for } h_1 < h \leq h_{i+1}: \quad Q = c_2 \, (h + a_2)^{b_2} \tag{20}$$

where: Q = discharge (m^3/s)
a_i, b_i, c_i = parameters
h_i, h_{i+1} = lower and upper water level for which the rating equation applies.

The coefficients a_1, b_1 and c_1 of the parabolic equation are determined by the least squares method. The shift parameter a_2 in the power equation is either input or determined by an adapted Johnson method. A computerized Johnson method (see for a description of the method e.g. WMO Operational Hydrology Report

2. normal fall method

In these methods the fall F between the water level at the discharge measuring site and a downstream station is considered as an additional parameter, to account for the effect of water surface slope on discharge. Both methods are based on the following equation:

$$Q_m = Q_r \left(\frac{F_m}{F_r}\right)^p \tag{23}$$

where Q_m - backwater affected discharge

Q_r - reference discharge

F_m - measured fall

F_r - reference fall

p - power, with: $0.4 \leq p \leq 0.6$

Constant fall method:

In this method the reference fall F_r is taken as a constant. A special case of the constant fall method is the unit fall method, where $F_r = 1$ m is applied.

In the computational procedure a value for F_r is assumed. Then a rating curve is fitted to the values:

$$Q_m = Q_r \left(\frac{F_r}{F_m}\right)^p \tag{24}$$

according to the standard procedure outlined in Chapter 7.2.3. The value for p is optimized between the boundaries 0.4 and 0.6 based on the least squares principle.

In the plot the fit of Q_r to the rating curve is shown, whereas in the error analysis the measured discharge Q_m is compared with the computed discharge according to eq.(23).

Normal fall method:

In this method the reference fall F_r is modelled as a function of the water level: $F_r = f(h)$. This function is represented by a parabola:

$$F_r = a_4 + b_4 h + c_4 h^2 \qquad (25)$$

valid for $h > h_{min}$ where h_{min} is a lower threshold of h above which the backwater correction is applied.

The normal fall method goes in two steps:

1. computation of the backwater free rating curve to represent the reference discharge Q_r
2. fitting of normal fall equation (14) to the reference falls.

Backwater free rating curve:

To isolate the backwater free data from the rest of the measurements a data flag = 2 has to be added to the data. This is done by selection of option <Adjust> from the screen and replacement of flag = 1 by 2 where applicable. Then to all backwater free data a rating curve is fitted analogous to the simple rating curve procedure described in Chapter 7.2.3. Note that sufficient backwater free data have to be available for a proper fit, else data have to be added (temporarily) to shape the rating curve.

Reference fall:

Next all remaining data above h_{min} with backwater effect are used to fit a parabola (25) to the reference fall computed as :

$$F_r = F_m \left(\frac{Q_r}{Q_m}\right)^{1/p} \qquad (26)$$

where; Q_r is the reference discharge $Q_r = f(h)$ computed from the backwater free stage discharge data. The parameter p is optimized between 0.4 and 0.6.

Example

Fitting of a simple rating curve to stage discharge measurements.The results are presented in Figure 19(a&b) and Table 7. A power type rating equation is used.

7.3 Validation of Rating Curve

7.3.1 General

In section 7.2 the following fitting procedures of stage discharge data is dealt with:

1. simple rating curve
2. rating curve with unsteady flow correction
3. rating curve with backwater corrections using:
   - constant fall, and
   - normal fall.

The option VALIDATION OF RATING CURVE covers the situation where a rating curve was developed previously and additional measurements become available. With HYMOS the validity of the existing rating curve for the new measurements can be evaluated.

7.3.2 Analysis

The analysis starts with the selection of the stage discharge measurements and the rating curve. In addition the number of data used for the fitting of the original rating curve and the standard error of the fit have to be entered for each water level interval, in which the rating curve has been splitted, to allow statistical tests.

Validation can be done in various ways:

1. examination of the goodness of fit from the stage discharge plot
2. examination of tabulated results:
   - difference between observed and computed discharge
   - time sequence of difference
   - relative difference
   - statistical test on fit.

HYMOS carries out a Student t-test on differences between the rating curve and the new measurements.

Let the number of data used to determine the rating curve be N and the number of new data N_1 and the percentage difference in both cases be denoted by respectively ΔQ and ΔQ_1 then the test statistics t becomes:

$$t = \frac{\overline{\Delta Q_1}}{a \cdot b} \tag{27}$$

with

$$a^2 = \frac{\Sigma(\Delta Q)^2 + \Sigma(\Delta Q_1 - \overline{\Delta Q_1})^2}{N + N_1 - 2} \tag{28}$$

$$b^2 = (N + N_1)/(N \cdot N_1) \tag{29}$$

7.4 Extrapolation of Rating Curve

7.4.1 General

The following procedures may be considered to extrapolate the stage discharge relation (ISO, 1982):

1. double logarithmic straight line extrapolation.
   If the hydraulic characteristics of the control section do not change beyond the measured range, the power type stage discharge relation is assumed to remain valid in the lower and upper end.
2. extension of the stage area (h-A) and stage velocity (h-v) curve beyond the measured range.
   This procedure is based on the observation that the h-v relation has little curvature under normal circumstances. The product of A and v can be used to obtain values of Q in the lower and upper end.
3. Extrapolation based on the Manning's or Chezy's equation for steady flow.

HYMOS includes graphical and numerical options to assist in the generation of synthetic stage discharge data in the extrapolation range by:

1. computation of cross sectional data
2. display of geometrical and hydraulic quantities in the measured range
3. computation of the same quantities in the extrapolated range from cross sectional data, slope and roughness.

7.5 Stage discharge Transformation

7.5.1 General

Water level series can be transformed in discharge series by use of a stage discharge relation of one of the following forms:

1. simple rating curve
2. rating curve with unsteady flow correction
3. rating curve with constant fall backwater correction
4. rating curve with normal fall backwater correction
5. general structure database
6. user structure equation
7. measuring structure relations

For a particular station the combination of water level(s) and discharge series, cross-sections and structure parameters required to transform stages into discharges can be stored in the database to facilitate the transformation at subsequent occasions.

8.0 DATA COMPILATION

8.1 (Dis-) Aggregation of Series

8.1.1 General

Aggregation of series implies the creation of a series with a larger time interval, by adding or averaging data of the series with the smaller interval: for example aggregation of series with an interval equal to one day to a series with an interval of one month, or from month to year, etc.. Disaggregation is the opposite process: by disaggregation, series with a smaller time interval are created. The various aggregation and dis-aggregation options are explained in the sections 8.1.2 and 8.1.3 respectively.

8.1.2 Aggregation

Basic and resulting series:

The original series. i.e. the series to be aggregated, is called basic series. The series with the aggregates is called the resulting series.

In aggregating series, two cases are distinguished:

1. for instantaneous observations the aggregated data are averages of the originals, e.g. discharges in m^3/s.
2. for accumulative observations the aggregated data are the sums of the originals, e.g. rainfall or discharges in mm.

8.1.3. Dis-aggregation

Basic and resulting series:

The original series i.e. the series to be dis-aggregated, is called basic series. The series with the dis-aggregates is called the resulting series.

In disaggregation of series following options exist:

1. for instanteneous observations
   - basic and resulting series data are equal
   - resulting series data are interpolated linearly between the mid point values of the basic series data;
2. for accumulative observations
   - resulting series data for fractions of the basic series data, derived by division by the number of resulting series intervals in a basic series interval,
   - as above, but additionally with a linear interpolation between the midpoints of the basic series intervals.es intervals.

8.2 Series Transformation

8.2.1 General

The series transformation option of HYMOS includes:

- algebraic transformation of series,
- transformation of non-equidistant series into equidistant series,
- creation of accumulative series, and
- error spreading.

These transformation options are dealt with in sections 8.2.2 to 8.2.5.

8.2.2 Algebraic transformations

The following algebraic transformations are possible to create a series Y by some function of series X_j, $j = 1,2,...$

1. linear equation

$$Y_i = C_1.X_{1,i} + C_2.X_{2,i} + C_3.X_{3,i} + C_4.X_{4,i} + C_5.X_{5,i} + C_6$$

2. multiplication $Y_i = X_{1,i} * X_{2,i}$

3. division $Y_i = X_{1,i} / X_{2,i}$

4. involution $Y_i = X_{1,i} \hat{} X_{2,i}$

5. natural logarithm $Y_i = \ln(X_i)$

6. common logarithm $Y_i = {}^{10}\log(X_i)$

7. exponential $Y_i = \exp(X_i)$

8. power of 10 $Y_i = 10 \hat{} X_i$

9. power $Y_i = X_i \hat{} C$

10. power of constant $Y_i = C \hat{} X_i$

where X_j = equidistant time series j

C_j = coefficients

In the application of the above transformations different start dates can be applied for each of the series.

8.2.3 Non-equidistant to equidistant series

Non-equidistant time series can be transformed into equidistant time series. Generally, the non-equidistant series may not fill all equidistant time steps. One can select one of the following options to solve this problem:

- zero:	the series values at intermediate time steps will be filled with zero's
- missing:	the series values at intermediate time steps will be filled with missing values
- linear:	the series values at intermediate time steps will be a linear interpolation between surrounding non-equidistant series observations
- equal to last:	the series values at intermediate time steps will be equal to the last observation, (i.e. block type filling in)

8.2.4 Accumulative series

Under the accumulative series option a series Y is created which is a continuous summation of a basic series X as follows:

$$Y_i = \sum_{j=1}^{i} X_j$$

8.3 Minimum and Maximum Search

This option of HYMOS comprises the computation of minimum, maximum and mean values for specific time periods.

Minimum, maximum and mean values for following time periods can be obtained:

- day
- month
- year
- period within the year

The results, i.e. minimum, mean and/or maximum series can be written to the data base or stored in a file.

8.4 Areal Rainfall

8.4.1 General

HYMOS offers a number of options to compute catchment rainfall from point rainfall data. The methods differ in the weights given to the point rainfall stations. Following weights are possible:

1. equal station weights,
2. user provided station weights,

3. weights according to Thiessen method, and
4. weights according to kriging method.

There are two ways to enter the codes and weights of the point rainfall stations:

1. the series are selected from the displayed series and the weights are entered manually (method 2) or are computed, or
2. the series codes and weights are read from a file, prepared at a previous occasion.

8.5 Kriging

8.5.1 General

HYMOS includes point and block kriging methods.

The point kriging option is used to:
- compute areal rainfall
- plotting of isolines (isohyets, isopotentials, etc.),
- design of measuring networks, and
- computation of rainfall station weights.

For computation of catchment rainfall and plotting of isohyets use is made of the kriging technique. For this a rectangular grid is placed over the catchment (=area of interest). Estimates of the rainfall at the grid points is obtained as a weighted average of the rainfall at surrounding stations. The station weights are determined by kriging and depend on the distances between the grid points and the rainfall stations on the one hand and the adopted covariance model on the other. The estimates for the grid points are expressed in a statistical way: the kriged value, which is the best estimate (unbiased and minimum variance), and the uncertainty in this value indicated by the standard deviation of the estimate. The technique can also be applied to other quantities with a spatial character. Well known is the application to derive groundwater level contours.

The block kriging method is used to compute, the best linear estimate and variance of rainfall in an area.

8.6 Computation of Evapotranspiration

8.6.1 General

HYMOS offers the following computational methods for potential evapotranspiration:

1. Penman method:
   - standard,
   - standard with FAO correction, and
   - dynamic,
2. FAO pan-evaporation method,
3. Christiansen method,
4. FAO radiation method,
5. Makkink radiation method,
6. Jensen-Haise method,
7. Blaney-Criddle method, and
8. Mass Transfer Method.

9 STATISTICAL ANALYSIS

9.1 Basic Statistics

9.1.1 General

The following statistics and distributions are included in this option:

- minimum value
- maximum value
- mean value
- median
- mode
- standard deviation
- skewness
- kurtosis
- empirical frequency distribution, and
- empirical cummulative frequency distribution

9.1.2 Selection of data

Data for statistical analysis can be entered in two ways:

- from HYMOS data base, or
- from file

If the data are read from the HYMOS data base following type of data can be considered:

- actual values
- annual minimum values, and
- annual maximum values.

9.2 Fitting Distributions

9.2.1 General

HYMOS includes the fitting of the commonly used theoretical frequency distributions:

1. Normal distribution
2. Log-normal distribution
3. Box-Cox transformation to normality
4. Pearson Type III or Gamma distribution
5. Raleigh distribution
6. Exponential distribution
7. General Pearson distribution
8. Log-Pearson Type III distribution
9. Extreme Type I or Gumbel distribution
10. Extreme Type II or Frechet distribution
11. Extreme Type III distribution
12. Goodrich/Weibull distribution
13. Pareto distribution, and
14. Peaks over Threshold (POT)-method for extremes (Pareto distribution)

For each distribution one can obtain:

- estimation of parameters,
- summary of observed and theoretical probabilities,
- goodness of fit-tests - binomial,
  - Kolmogorov-Smirnov,
  - Chi-squares,
- computation of extreme values for specific return periods, either related to probability of non-exceedance or exceedance, and
- plot of distribution function with 95% confidence limits (optional).

Table 8 and Figure 20 shows the results in tabular and graphical form respectively.

9.3 Statistical Tables

9.3.1 General

Variates, probabilities of non-exceedance and return periods can be computed for the following frequency distributions:

1. Normal distribution
2. Log-normal distribution

9.2 Fitting Distributions

9.2.1 General

HYMOS includes the fitting of the commonly used theoretical frequency distributions:

1. Normal distribution
2. Log-normal distribution
3. Box-Cox transformation to normality
4. Pearson Type III or Gamma distribution
5. Raleigh distribution
6. Exponential distribution
7. General Pearson distribution
8. Log-Pearson Type III distribution
9. Extreme Type I or Gumbel distribution
10. Extreme Type II or Frechet distribution
11. Extreme Type III distribution
12. Goodrich/Weibull distribution
13. Pareto distribution, and
14. Peaks over Threshold (POT)-method for extremes (Pareto distribution)

For each distribution one can obtain:

- estimation of parameters,
- summary of observed and theoretical probabilities,
- goodness of fit-tests - binomial,
  - Kolmogorov-Smirnov,
  - Chi-squares,
- computation of extreme values for specific return periods, either related to probability of non-exceedance or exceedance, and
- plot of distribution function with 95% confidence limits (optional).

Table 8 and Figure 20 shows the results in tabular and graphical form respectively.

9.3 Statistical Tables

9.3.1 General

Variates, probabilities of non-exceedance and return periods can be computed for the following frequency distributions:

1. Normal distribution
2. Log-normal distribution

Examples of frequency and duration curves are shown in figure 21 and table 9.

10. Time series analysis

10.1 Introduction

Time series analysis includes the execution of following types of analysis:

1. Correlation analysis
2. Spectral analysis
3. Range analysis
4. Run analysis
5. Storage analysis

The analyses are described in the next sections.

IMPORTANT; note that the time series which are investigated should not contain missing values:

10.2 Correlation analysis

10.2.1 General

Correlation analysis covers the computation of:

1. auto-covariance function
2. auto-correlation function
3. cross-covariance function
4. cross-correlation function

HYMOS produces tabular and graphical presentations of these functions.

10.2.2 Auto-covariance and auto-correlation functions

For the series x_i, $i=1,N$ the auto-covariance function $c_{xx}(k)$, $k = 0, L_{max}$ is computed as follows:

$$c_{xx}(k) = \frac{1}{N} \sum_{i=1}^{N-k} (x_i - m_x)(x_{i+k} - m_x) \qquad (30)$$

where:

m_x = average of x_i , $i=1,N$.

k = time-lag in time units equal to the time interval

L_{max} = maximum lag

The auto-correlation function $r_{xx}(k)$ is determined from:

$$r_{xx}(k) = c_{xx}(k)/c_{xx}(o) \qquad (31)$$

The 95% tolerance or confidence limits for zero correlation according to Siddiqui (see Yevjevich (1972)) are computed from:

$$CL_p(k) = \frac{-1}{N-k+1} +1.96 \frac{(N-k-1)}{(N-k+1)} \sqrt{\frac{1}{N-k}} \qquad (32)$$

$$CL_n(k) = \frac{-1}{N-k+1} -1.96 \frac{(N-k-1)}{(N-k+1)} \sqrt{\frac{1}{N-k}} \qquad (33)$$

where:

$CL_p(k)$ = upper confidence limit for zero correlation at lag k

$CL_n(k)$ = lower confidence limit for zero correlation at lag k

An example of an auto-correlogram of monthly rainfall, shown in figure 22 is presented in Table 10 and figure 23.

10.2.3 Cross-covariance and cross-correlation functions

The cross-covariance functions $c_{xy}(k)$,and $c_{yx}(k)$=0, L_{max} are computed as follows:

$$c_{xy}(k) = \frac{1}{N} \sum_{i=1}^{N-k} (x_i - m_x)(y_{i+k} - m_y) \qquad (34)$$

$$c_{yx}(k) = \frac{1}{N} \sum_{i=1}^{N-k} (x_{i+k} - m_x)(y_i - m_y) \qquad (35)$$

where:

m_x = average of x_i, i=1, N

m_y = average of y_i, i=1, N.

The cross-correlation functions $r_{xy}(k)$ and $r_{yx}(k)$ are estimated from :

$$r_{xy}(k) = c_{xy}(k)/(s_x . s_y) \qquad (36)$$

$$r_{yx}(k) = c_{yx}(k)/(s_x . s_y) \qquad (37)$$

where:

s_x = standard deviation of x_i, $i=1,N$

s_y = standard deviation of y_i, $i=1,N$

10.3 Spectral analysis

The smoothed auto-spectral estimate $C_{xx}(f)$, for f=0,..,1/2 is calculated from:

$$C_{xx}(f) = 2 \{C_{xx}(0)+2 \sum_{k=1}^{M-1} C_{xx}(k)w(k) \cos(2\pi fk)\} \qquad (38)$$

where: f = frequency in cycles per time interval, computed at spacings $1/(2N_f)$, where N_f is 2 to 3 times M

N_f = number of frequency points

$C_{xx}(k)$ = autocovariance function at lag k

M = truncation point or maximum lag of the autocovariance function used to estimate the autospectrum; clearly M is conditioned by: $M \le L_{max}$

w(k) = window function

Following window w(k) for k=1, M-1 according to Tukey is used to smoothen the spectral estimate:

$$w(k) = \frac{1}{2}\ (1+\cos(\frac{\pi k}{M})) \qquad (39)$$

The band B and number of degrees of freedom are given by:

$$B = 4/(3M) \qquad (40)$$

$$N_f = 8N(3M) \qquad (41)$$

The logarithm of the auto-spectrum is computed by:

$$C_{log}(f) = \log^{10} C_{xx}\ (j) \qquad (42)$$

In the results $C_{log}(f)$ will be set to -100 if $C_{xx}(f) \le 0$.

The spectral density function follows from:

$$R_{xx}(f) = \frac{C_{xx}(f)}{C_{xx}(o)} = 2\{1+2 \sum_{K=1}^{M-1} r_{xx}(K)\, W\,(K) \mathrm{Cos}\,(2\pi f K)\} \qquad (43)$$

An example is presented in Table 11 and Figure 24, where the spectral density function of monthly rainfall given in Figure 22 is shown.

10.4 Range analysis

In Figure 25 a definition sketch of the following range related quantities is given :

- adjusted surplus ${}_aS_N^+$
- adjusted deficit ${}_aS_N^-$
- adjusted range ${}_aR_N$, and
- rescaled adjusted range ${}_aR_N^*$.

The quantities are computed from the accumulative departures from the mean S_i for i = 0, N and with $S_o = 0$;

$$S_i = \sum_{j=o}^{i} (x_j - m_x) c_f \qquad (44)$$

where: m_x = average of x(i), i = 1, N

c_f = conversion factor (time units per time interval) to transfer intensities into volumes

It follows for:

- Surplus ${}_aS_N^+$:

$$ {}_aS_N^+ = \max\,(S_0, S_1 \ldots\ldots, S_N) \qquad (45)$$

- Deficit ${}_aS_N^-$:

$$ {}_aS_N^- = \min\,(S_0, S_1, \ldots\ldots S_N) \qquad (46)$$

- Adjusted range ${}_aR_N$:

$$ {}_{\alpha}R_N = {}_{\alpha}S_N^+ - {}_{\alpha}S_N^- \tag{47} $$

- Rescaled adjusted range ${}_{\alpha}R_N^*$:

$$ {}_{\alpha}R_N^* = {}_{\alpha}R_N / (S_x . c_f) \tag{48} $$

where : S_x = standard deviation of x_i, $i = 1, N$

10.5 Run analysis

A definition sketch for run analysis is presented in Figure 26 .

Up and down crossing and runs:
Let x_c be a crossing level then an up crossing is defined by:

$$ x_{i+1} \geq x_c \text{ and } x_i < x_c \tag{49} $$

and a down crossing by:

$$ x_{i+1} < x_c \text{ and } x_i \geq x_c \tag{50} $$

A run is an excursion above or below the level x_c, i.e. bounded by an up crossing and a down crossing or a down crossing and an upcrossing. Note: HYMOS also interprets as runs the first and last excursion above or below level x_c, which are only bounded by an upcrossing or a downcrossing; these runs are incomplete.

Run length

With respect to run length, the following distinction has to be made:

- positive run length RL^+,
- negative run length RL^-, and
- total run length, i.e. successive pair of $RL^+ + RL^-$

RL^+ is the time space between an upcrossing and downcrossing and RL^- is the time span between a down crossing and an up crossing given as a number of time intervals.

Run sum

The positive and negative runsums RS^+ and RS^-, respectively, are computed from :

$$RS^+ = \sum_{i=j}^{k} (x_i - x_c)C_f \qquad (51)$$

where,

j	=	location of an up crossing
k	=	location of the next down crossing
C_f	=	conversion factor (= time units per time interval) to transfer intensities into volumes

$$RS^- = \sum_{i=k}^{m} (x_c - x_i)\, C_f \qquad (52)$$

where k = location of the down crossing
m = location of the next up crossing

10.6 Storage Analysis

By the sequent peak algorithm water shortage or equivalently storage requirements without running dry are computed for various draft levels from the reservoir. The procedure is a computerized variant of the well known graphical Rippl technique. The algorithm considers the following sequence of storages:

$$S_i = S_{i-1} + (x_i - D_x)C_f \qquad i=1,\ 2N \qquad (53)$$

where

x_i	=	inflow
D_x	=	$D_L . m_x$
m_x	=	average of x_i, $i=1,\ N$
D_L	=	draft level as a fraction of m_x
C_f	=	multiplier to convert intensities into volumes (time units per time interval)

The local maximum of S_i larger then the preceding maximum is sought. Let the locations be k2 and k1 respectively with k2 > k1. Then the largest positive difference between S_{k1} and S_i, i = k1, k2 is determined, which is the local range. This procedure is continued to i = 2N, i.e. two times the length of the actual series x_i. In this way initial effects are eliminated.

11.0 REPORTING AND RETRIEVAL

11.1 Reports

11.1.1 General

The option reports includes the ready for publication preparation of station reports:

- station characteristics
- series characteristics
- current metering data and rating parameters
- station reports

The various options are dealt with in the Chapters 11.1.2 to 11.1.4 respectively.

Note:start with setting of the <Selection> switch on and set the output parameters by selecting <Options>.

11.1.2 Print of station characteristics

Following output can be obtained:

- station characteristics, and
- station histories

Station characteristics:
For printing of station characteristics following steps have to be taken:

- Select <Station> from the main options.
- Select <Characteristics> from the next menu.,

If the selection switch is off the characteristics of all station of the 'current data base' will be printed.

If the selection switch is on selection by one of the following items can be made:

. Station Code	. Latitude
. Station name	. Longitude
. River	. Altitude
. District	. Catchment area, and
. Country	. Agency

For the selected item one or a range of names/values may be specified by selecting respectively Equals or Range and by entering subsequently the name(s)/value(s).

Station histories:
For printing of station histories following steps have to be

taken:

- Select <Station> from the main options
- Select <History> from the next menu

If the selection switch is off the histories of all stations of the 'current data base' will be printed.

If the selection switch is on a selection on station code can be made. One or a range of station codes may be specified by selecting respectively <Equals> or <Range> and by entering subsequently the station code(s).

11.1.3 Print of series characteristics and data

Following output can be obtained:

- series characteristics
- current-metering data
- stage-discharge parameters
- sediment-discharge parameters

Series characteristics:

Series characteristics cover: series code, time and basic time interval, time shift, observation type and start and end date of data availability. For printing of these characteristics following steps have to be taken.

- Select <Series> from the main HYMPRN options.
- Select <Characteristics> from the next menu

If the selection switch is off the characteristics of all series of the 'current data base' will be printed.

If the selection switch is on a selection on one of the following items can be made:

. Station code	. Interval, and
. Type	. Period

For the selected item one or a range of names/values may be specified by selecting respectively <Equals> or <Range> and by entering subsequently the name(s)/value(s).

Current metering data:

For printing of current metering data following steps have to be taken:

- Select <Series> from the main options.
- Select <Current-metering data> from the next menu
- If the selection switch is off all available current metering data for the 'current data base' will be printed.
- If the selection switch is on a selection on one of

the following items can be made:

- Station code
- Date
- Number
- Gauge zero
- Water level
- Discharge
- Gradient
- Width
- Wetted perimeter
- Area
- Velocity, and
- Hydraulic radius

For the selected item one or a range of names/values may be specified by selecting respectively <Equals> or <Range> and by entering subsequently the name(s)/value(s).

Stage-discharge parameters:
For printing of discharge rating curve parameters following steps have to be taken:

- Select <Series> from the main options.
- Select <Discharge rating parameters> from the next menu.

If the selection switch is off all available discharge rating parameters for the 'current data base' will be printed.
If the selection switch is on a selection on one of the following items can be made:

- Station code, and
- Date from-to (validity of the rating curve)

For the selected item one or a range of names/dates may be specified by selecting respectively <Equals> or (Range) and by entering subsequently the name(s)/dates. The start and end date should be entered as <YYYYMMDDyyyymmdd> without space in between the year, month and day values and the start and end date.

Sediment rating parameters:
For printing of sediment rating curve parameters following steps have to be taken:

- Select <Series> from the main options.
- Select <Sediment-rating parameters > from the next menu
- If the selection switch is off all available sediment rating parameters for the 'current data base' will be printed.
- If the selection switch is on a selection on one of the following items can be made:

. Station code, and

. Date from-to (validity of the rating curve)

For the selected item one or a range of names/dates may be specified by selecting respectively <Equals> or (Range) and by entering subsequently the name(s)/dates. The start and end date should be entered as <YYYYMMDDyyyymmdd> without space in between the year, month and day values and the start and end date.

11.1.4 Print of station reports

Following reports station wise can be obtained:

1. Station characteristics, and/or
2. Station histories, and
3. Time series data.

An example in shown in table 12 .

For printing of the reports following steps have to be taken:

- Select <Report> from the main options
- Select a station from the next menu
- Select time interval of time series to be printed in the report.
- Select the data types from the available series
- Specify requirement of:
  - . Station characteristics
  - . Station history

- Enter the start data.
- Enter the number of months/years to be printed

11.2 Mixed tables

Tables of equidistant time series of mixed quantities can be presented in a coherent column user-friendly way, in a form ready for reports. HYMOS offers the following options:

1. Free text for table headers and foot-notes
2. Free text for column headers.
3. Decimal setting and rounding optional.
4. Coding of data:
   - original
   - filled in
   - corrected
5. Basic statistics of column-wise including (optional):
   - number of data
   - minimum and maximum
   - sum

- mean
- standard deviation
- coefficient of variation

An example of this option is shown in Table 13 .

11.3 Graphics

11.3.1 General

The graphics options cover the drawing of functions and of isolines with many facilities for reporting.
To produce figures, following input files are required:

1. a picture file, including information about the lay-out of the figure and about the functions to be plotted. The user creates these files with the extension .PCT.
2. data files, contain information (function values) to be drawn. The user creates these files with extension .MAT. More than one data file can be used.

Furthermore, information has to be added on the plot configuration, plot device, etc.

The set up of the graphics option is modularly. The various parts will be described separately. Through these parts so-called "tasks" are created. The .PCT files is a collection of tasks which describe a drawing and also define the functions to be plotted.

Following main options are available:

1. Data: to create data (.MAT) -files
2. Pre-processor: to create picture (.PCT) -files, and
3. Draw: to make a plot
4. Edit: to edit data (.MAT) - files

   In the following section some of the options are explained.

11.3.2 Graphics options

The graphics option comprises a number of elements applied when using computer graphics.

Co-ordinates:

The available, data, such as measured values and computations are positioned by co-ordinates when displayed in graphics. A few examples: water levels, discharge (both time functions), an area with measuring points, an area with isoline.

Window:

A window is placed over the available data in order to indicate which data is to be presented. Only data within the frame of the window are drawn.

Viewport:

A rectangular area on the screen or on paper in which the picture is drawn is called a viewport. More viewports can exist, but also one only, namely the whole screen or paper. A window is displayed within a viewport by means of scaling. Through translation or rotation, a viewport is displaced along the screen or the paper.

Clipping and blanking:

The part of the data, not covered by the window is left out of consideration. This process is called 'clipping'.

It is also possible that a viewport covers another viewport, or a part of it, resulting in covering of a part of the underlying viewport. Also it can be indicated by means of a contour that a part of the picture within a viewport is unwanted. The above described covering is called 'blanking'.

The area that is covered can either remain empty, or not. It can also be used to insert information later on, such as text. The contour is often placed in position by means of a polygon.

Polygon:

A polygon is a set of x, y-values, connected to each other by straight line parts. Polygons can be defined from two data array, or they are directly available in a data file.

Example are shown in figure 27.

11.4 Retrieval and transfer

11.4.1 General

HYMOS includes following options for data retrieval and transfer:

1. retrieval of data for external usage,
2. transfer to internal HYMOS data base, and
3. transfer to external HYMOS data base.

The options are discussed in the sections 11.4.2 and 11.4.3.

11.4.2 Transfer to user files

The data stored in the data base can be retrieved for transfer to file. This covers following data:

1. equidistant time series,
2. non-equidistant data, including:
   - station and history data,
   - time series,
   - current metering data,
   - stage discharge parameters, and
   - series relation parameters.

Station/series code, start and end date of series, field length specification and user file name are generally required to execute the transfer.

11.4.3 Transfer to HYMOS data base

In the transfer to a HYMOS data base, two cases are distinguished:

1. transfer to an internal data base, and
2. transfer to an external data base.

Transfer to internal data base:

This refers to transfer of series to another HYMOS data base on the same system or in the user network, e.g. from a temporary data base to a final one storing only validated and corrected data. The required input includes:

- series code,
- data drive of the receiving data base,
- name and extension of the receiving data base, and
- start and end date

For this transfer it is necessary that the characteristics of the series to be transferred are available on both data bases.

Transfer to external data base:

This refers to transfer of series to another HYMOS data base on a different system. HYMOS creates a data file with the proper read instructions for easy load in the other data base. The input includes:

- series code,
- field length and number of decimals,
- start and end date of the period to be transferred, and
- name of file the data will be (temporarily) stored.

REFERENCES :

1. Data Systems and Management:

   · An introduction to System Analysis and Design, 1980 by Atton R Kindred.

2. Database technology A software engineering Approach, 1988, John G. Hugue.

3. Guidelines for computerized data processing in operational hydrology and land and water management;Joint FAO/WMO publication ; WMO No. 634 (1985)

4. HYMOS user manual, Delft Hyraulics.

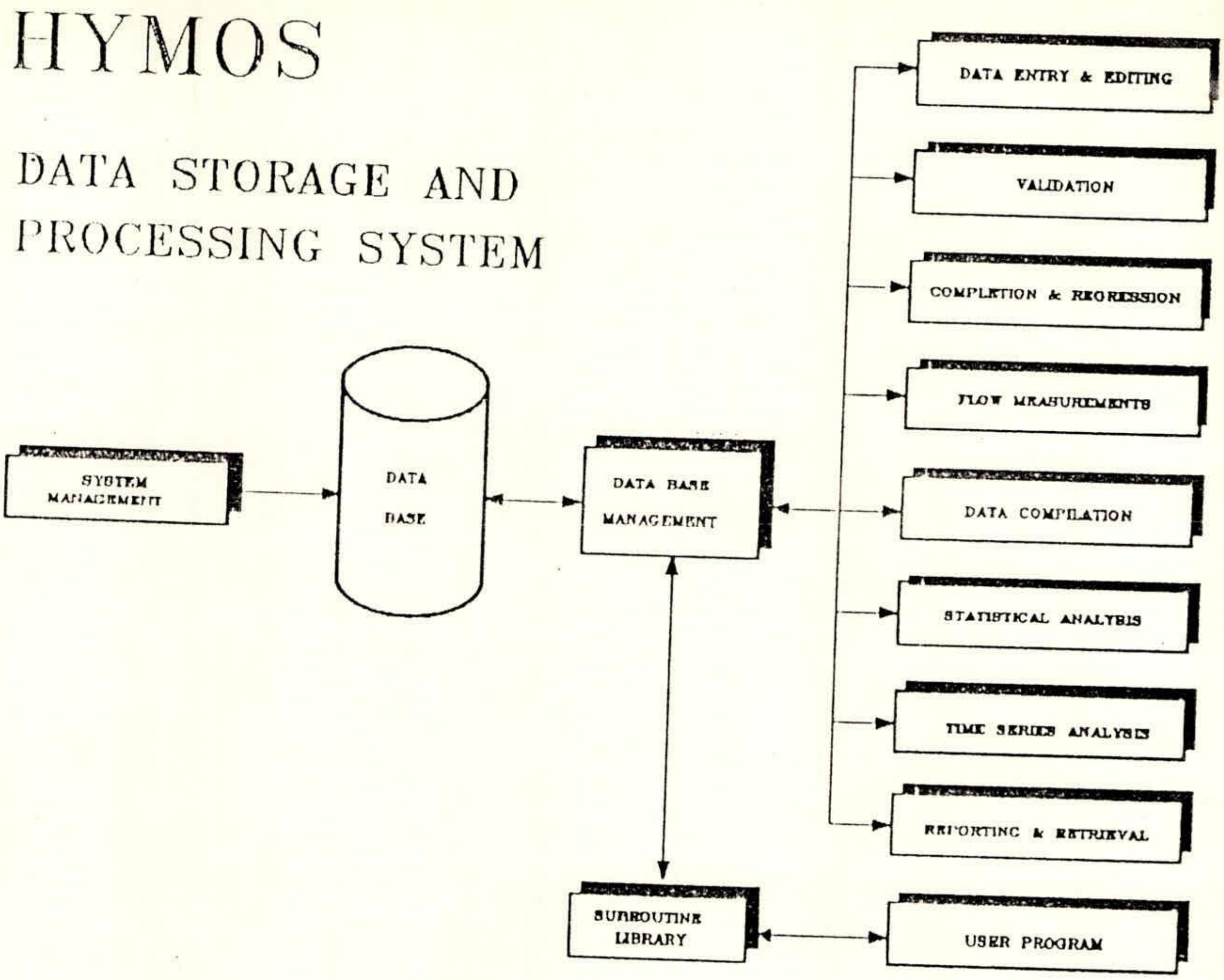


Figure 1. Structure of HYMOS.

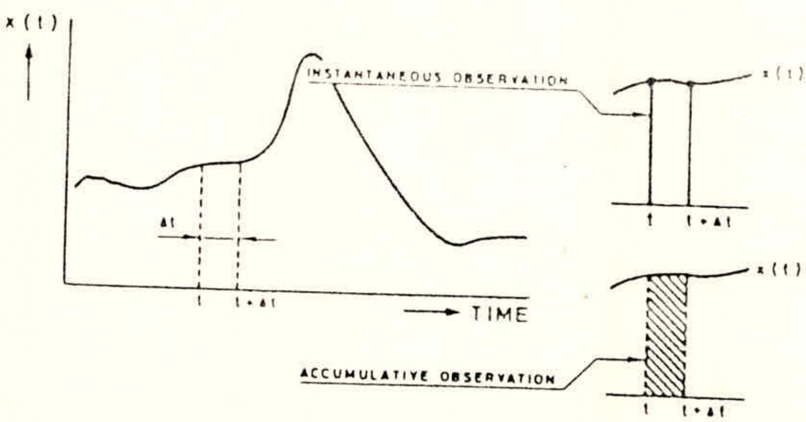


Figure 2 Instantaneous and accumulative observations

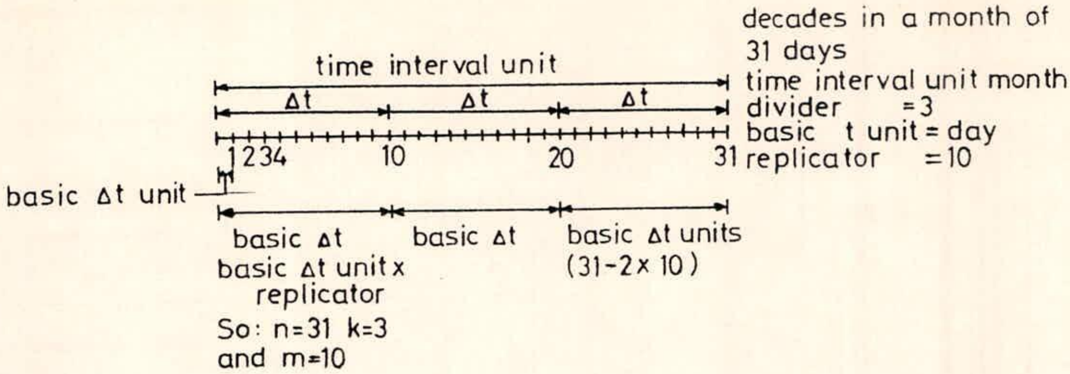


Fig. 3- Explanation of basic time interval

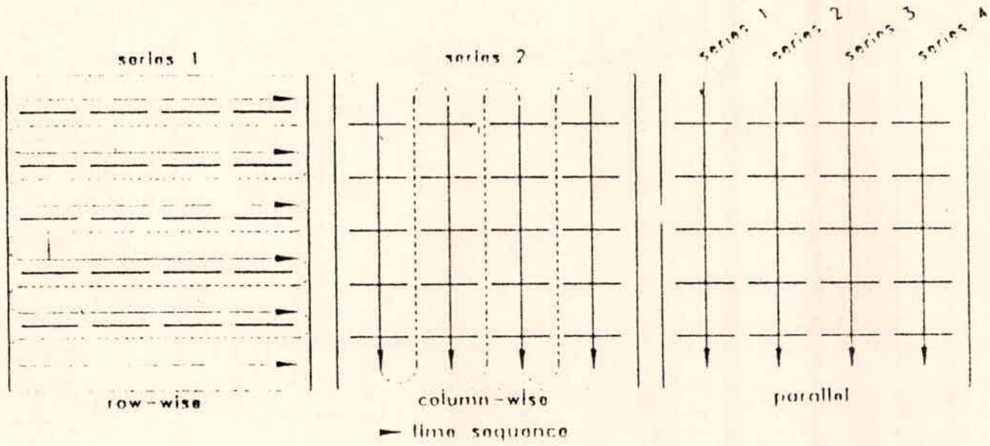


Figure 4 Row-wise, column-wise and parallel organized data files

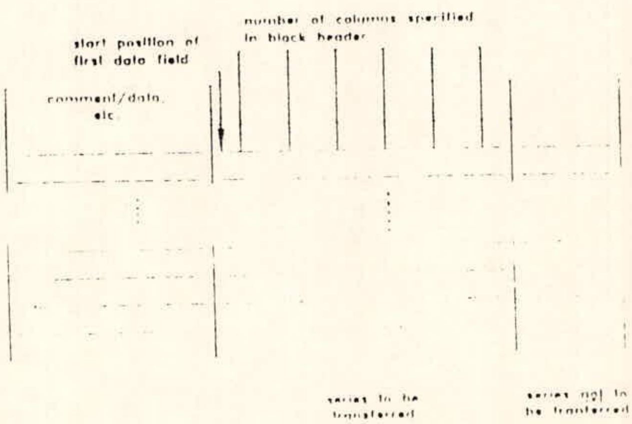


Figure 5 Layout of FIXEd format row-wise time sequential data file

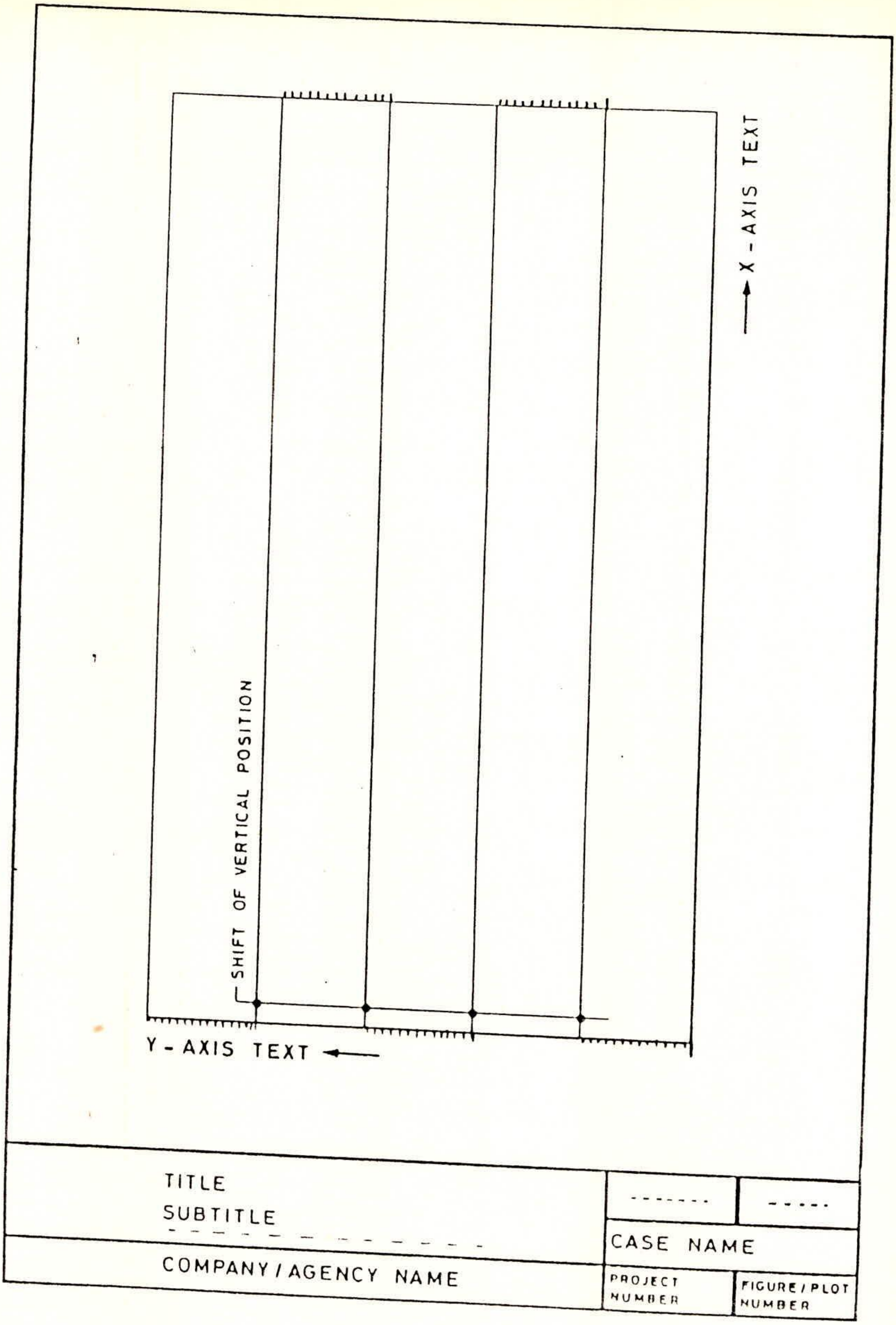


Figure 6 General layout of a time series graph in A4 cadre

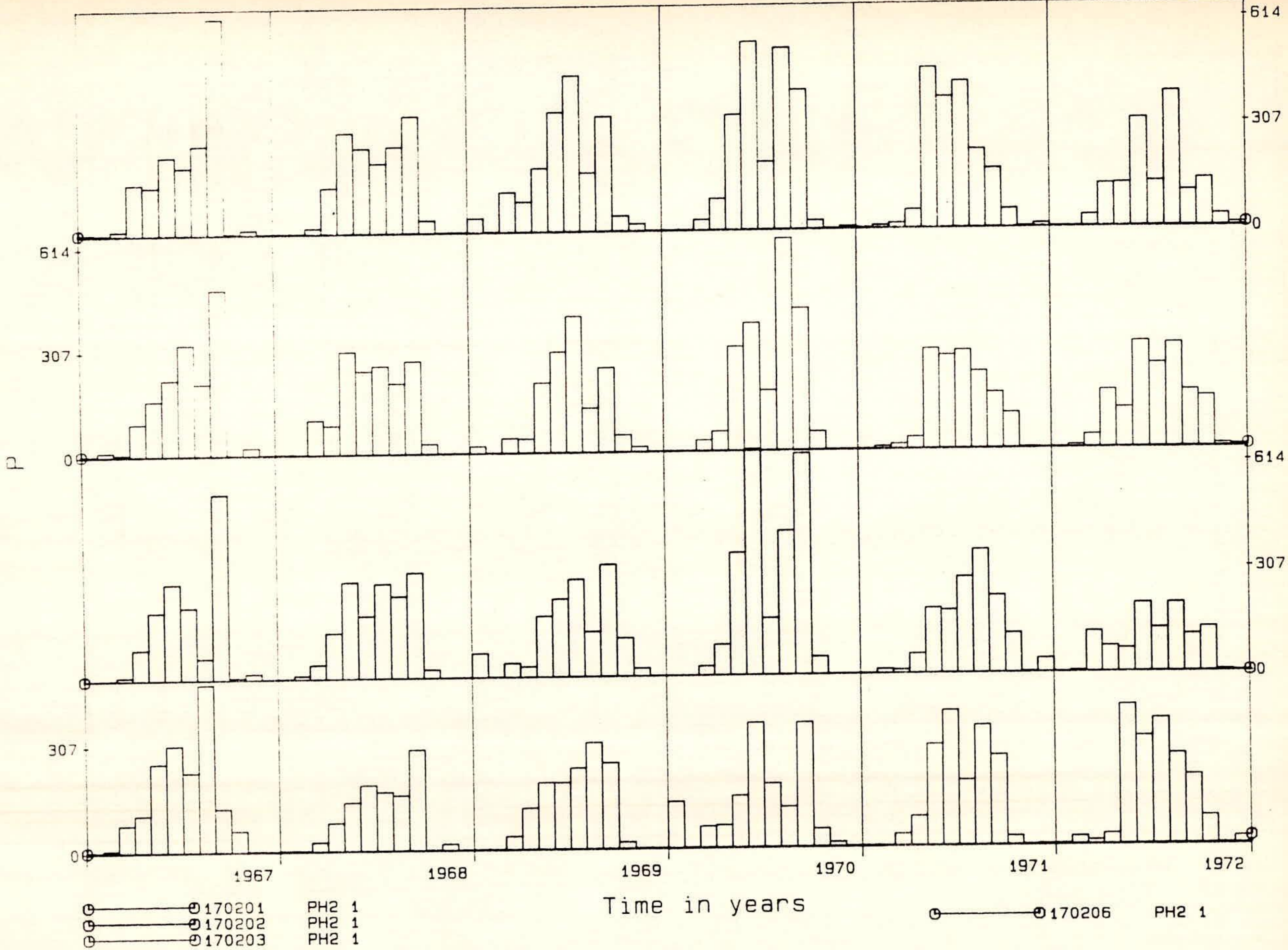


FIG. 7. Example of time series

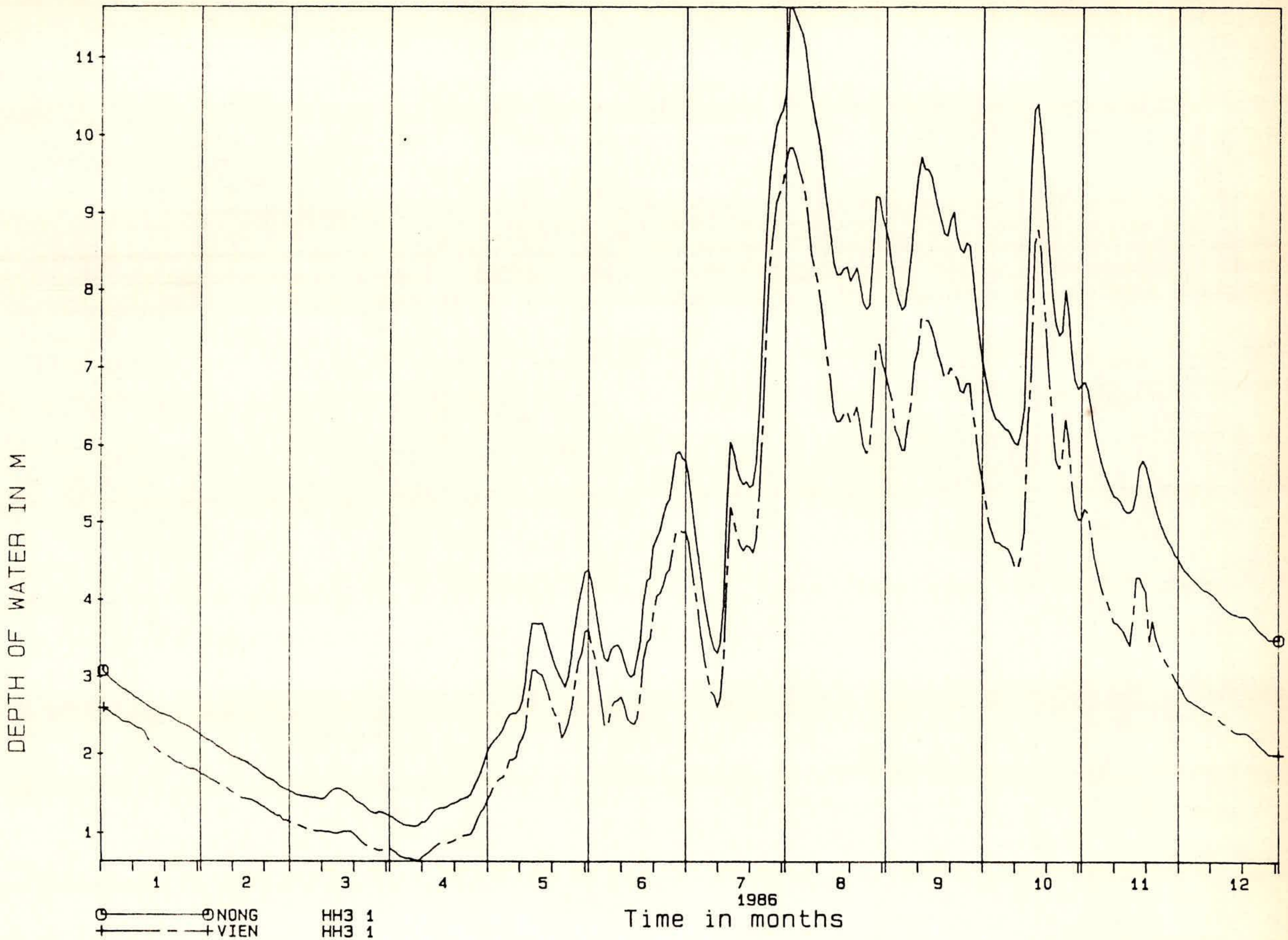


FIG. 8 Example of time series

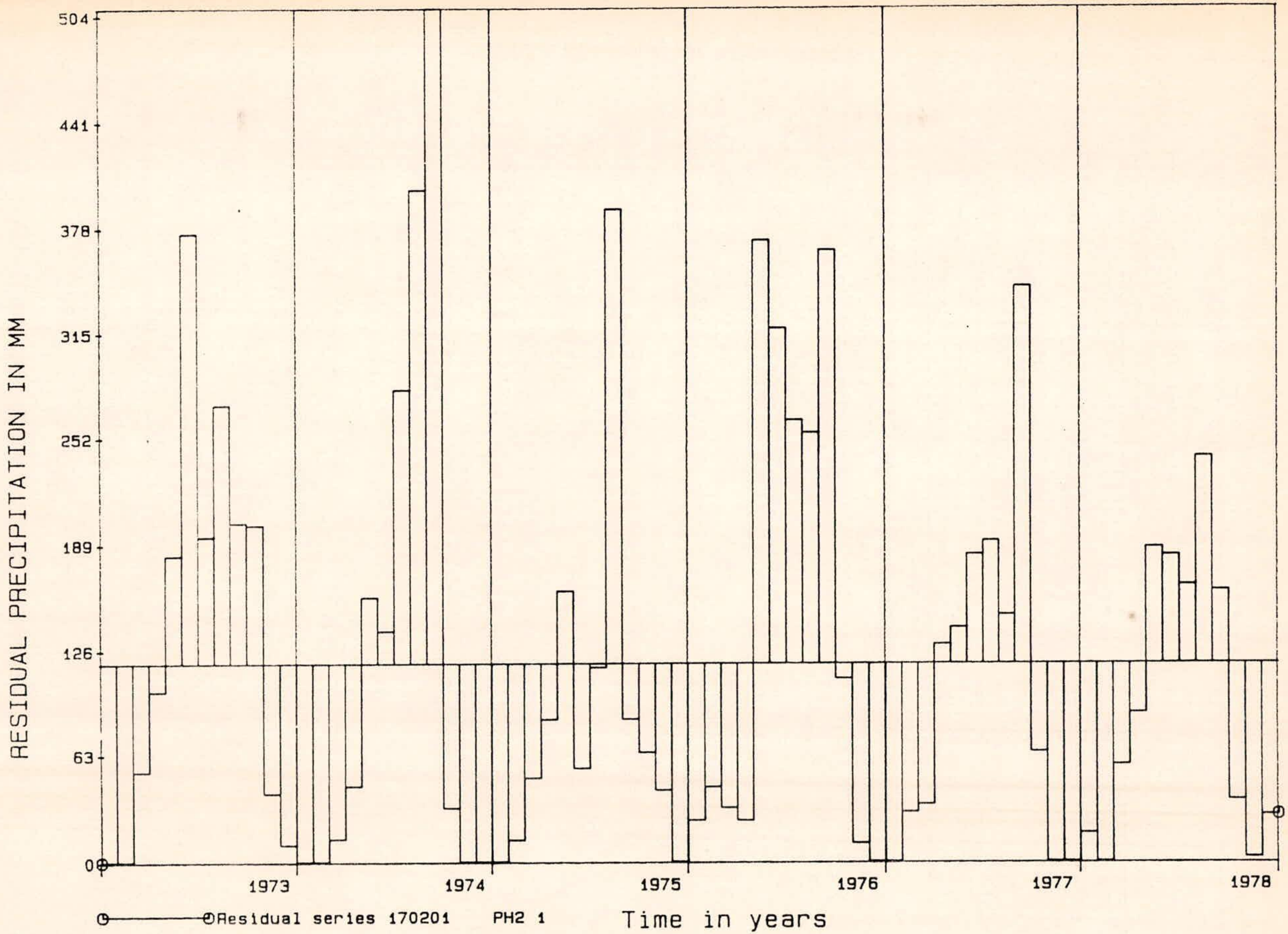


FIG. 9 Example of residual series

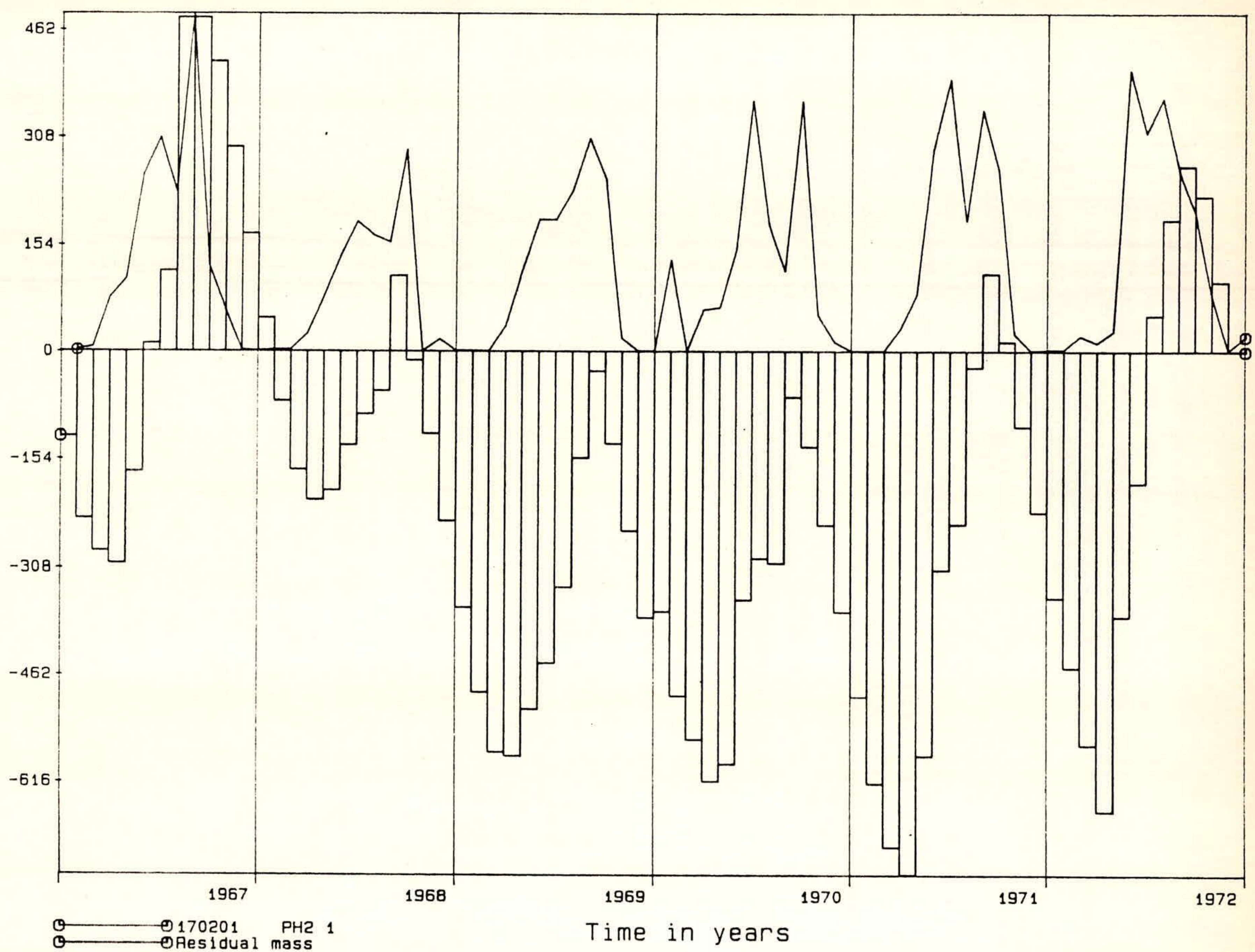


FIG. 10 Example of residual mass curve

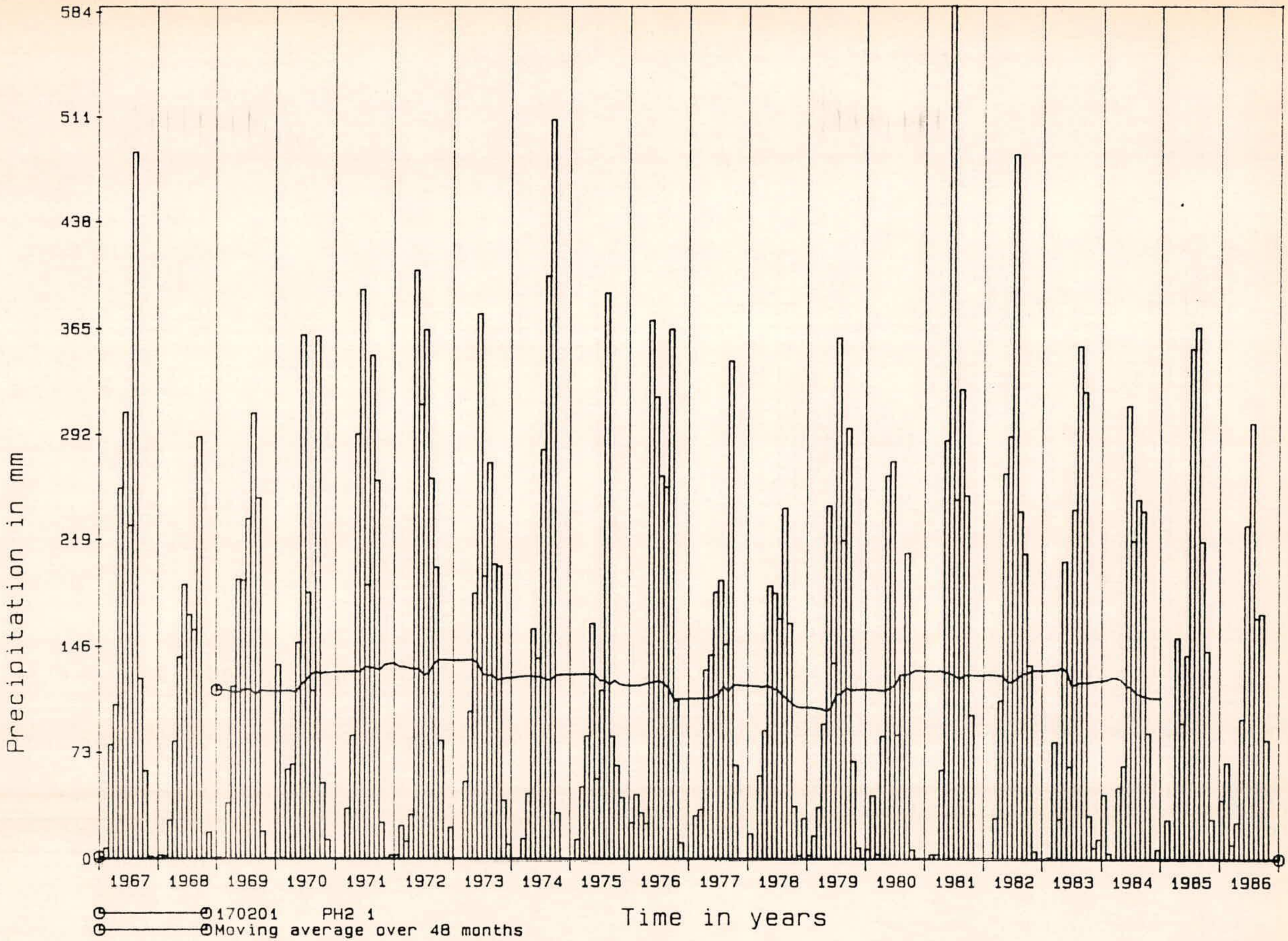


FIG. 11 Example of moving average curve

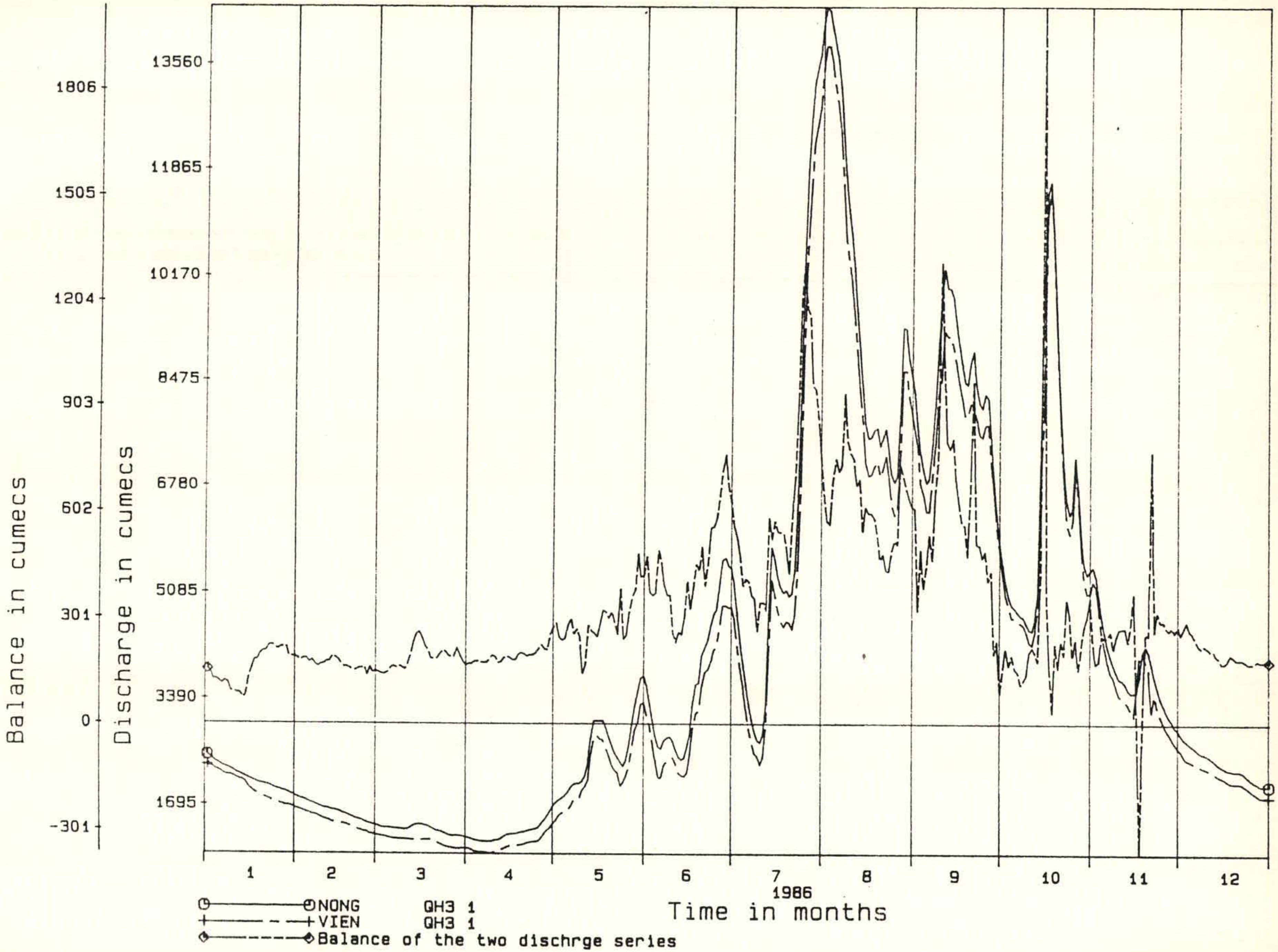


FIG. 12 Example of balance graph

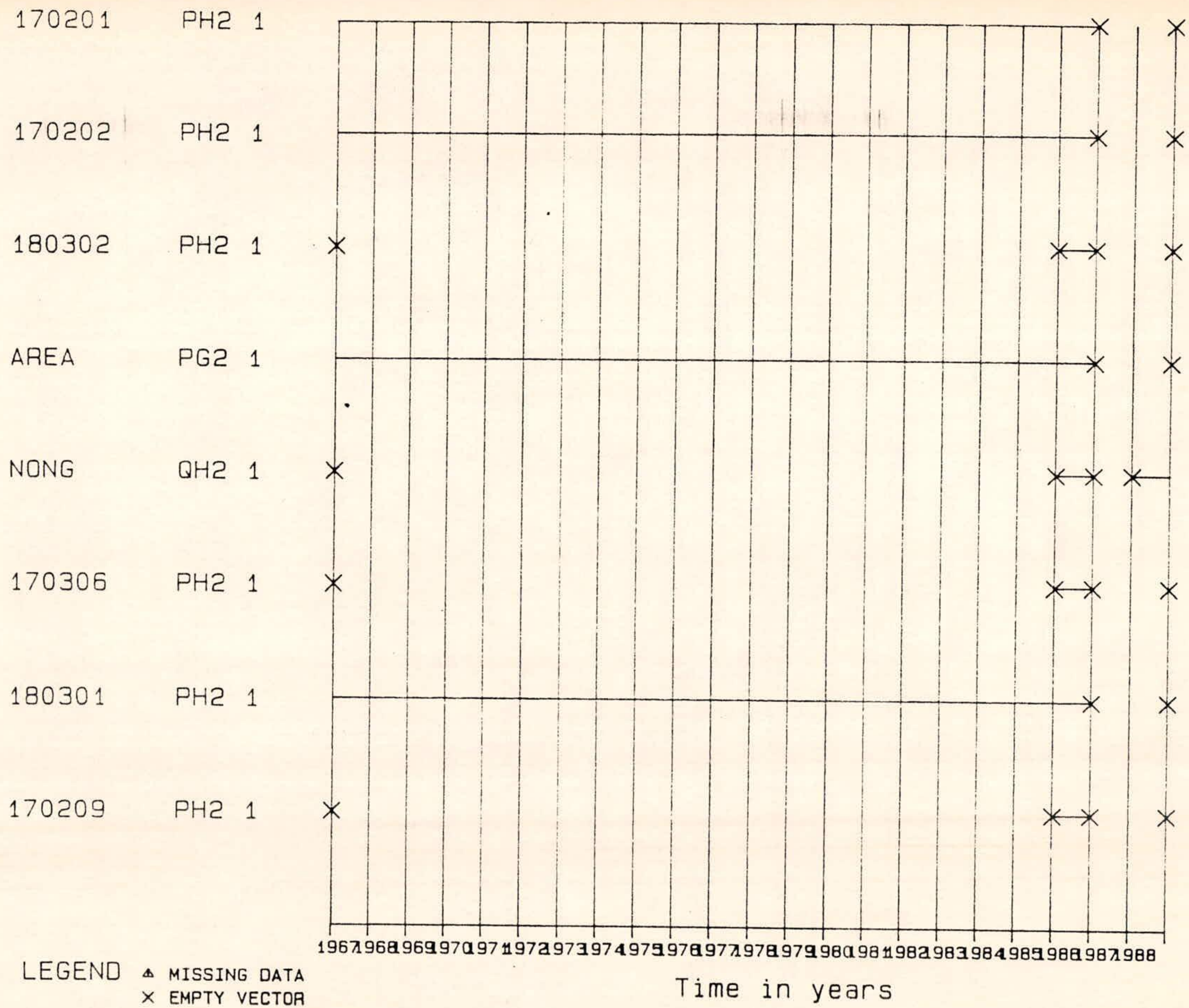


FIG. 13 Example of bar chart of series availability

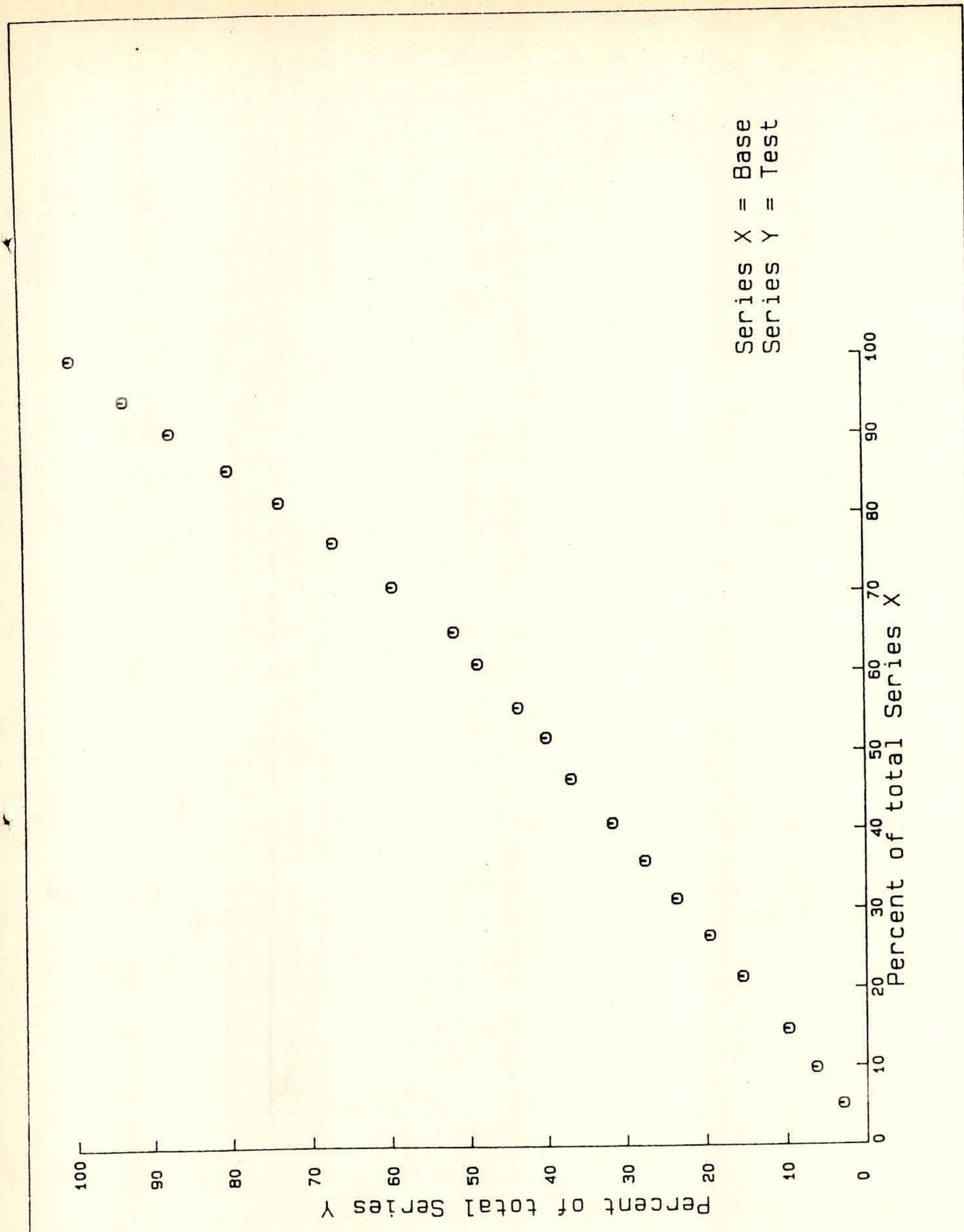
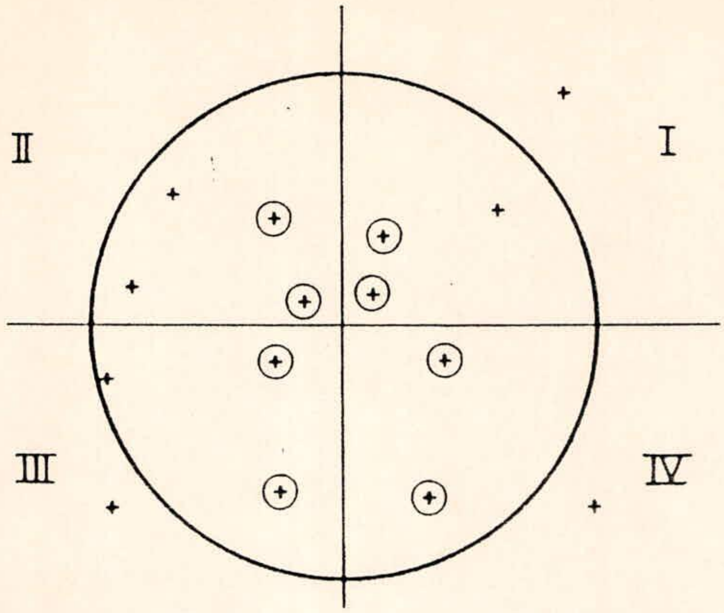


FIG. 14 Example of double mass analysis

DOUBLE MASS ANALYSIS
BASE - 170201, 170202, 170203, 170206
TEST - 180301

REPORT	
1.1.92	FIG. 19

NATIONAL INSTITUTE OF HYDROLOGY



Legend:

\+ - neighbour station

(+) - selected neighbour station

I, etc. - quadrant

Figure 15 Definition sketch

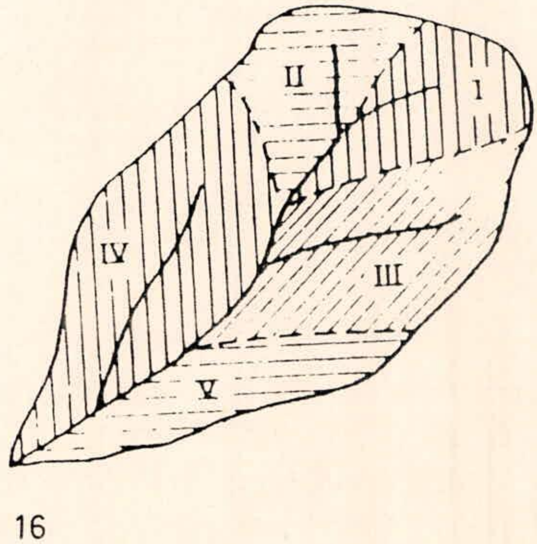


FIG. 16

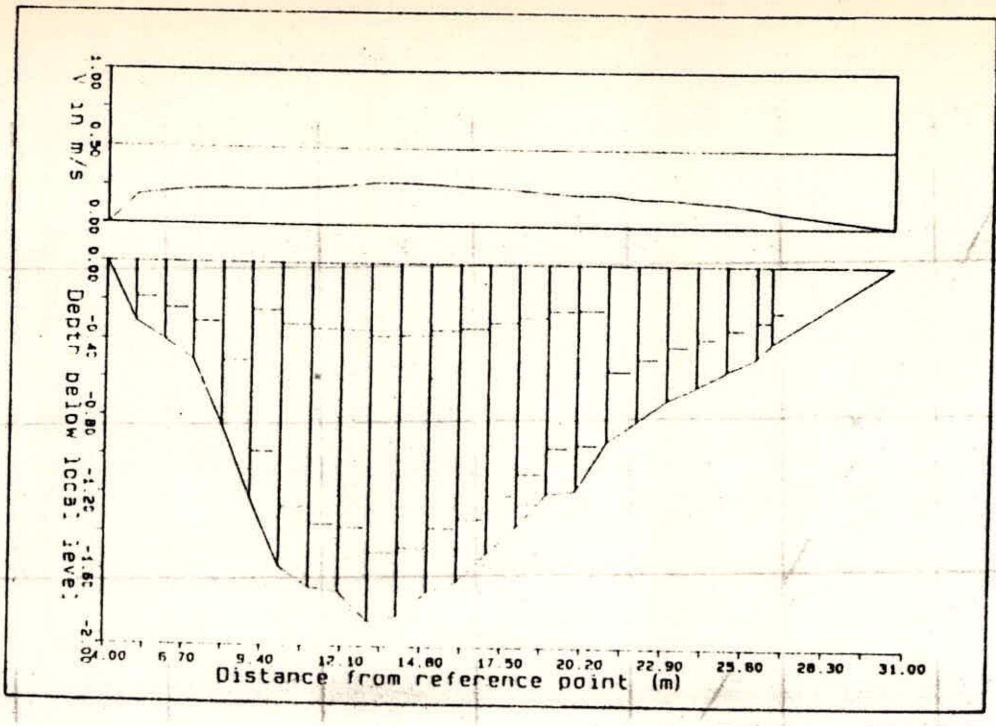


Figure 17 Flow velocities in the cross-section.

Figure 18 Estimation of the adjustment factor $1/SV_w$

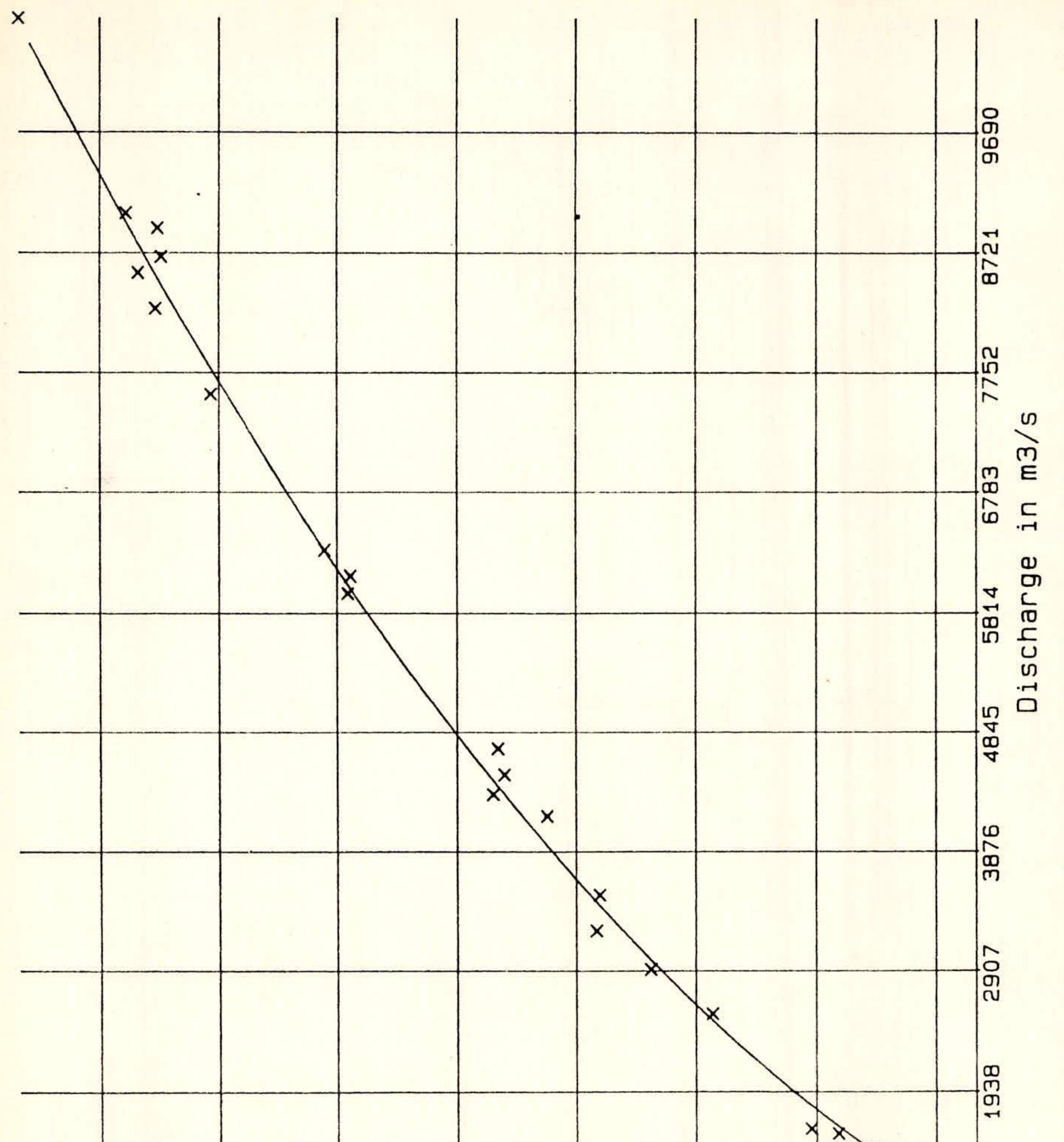


FIG.19(a) Fitting of stage discharge data by a simple rating curve

Fitting Simple Rating Curve GD Site at Haridwar (hypothetical)		Report
National Institute of Hydrology		FIG.

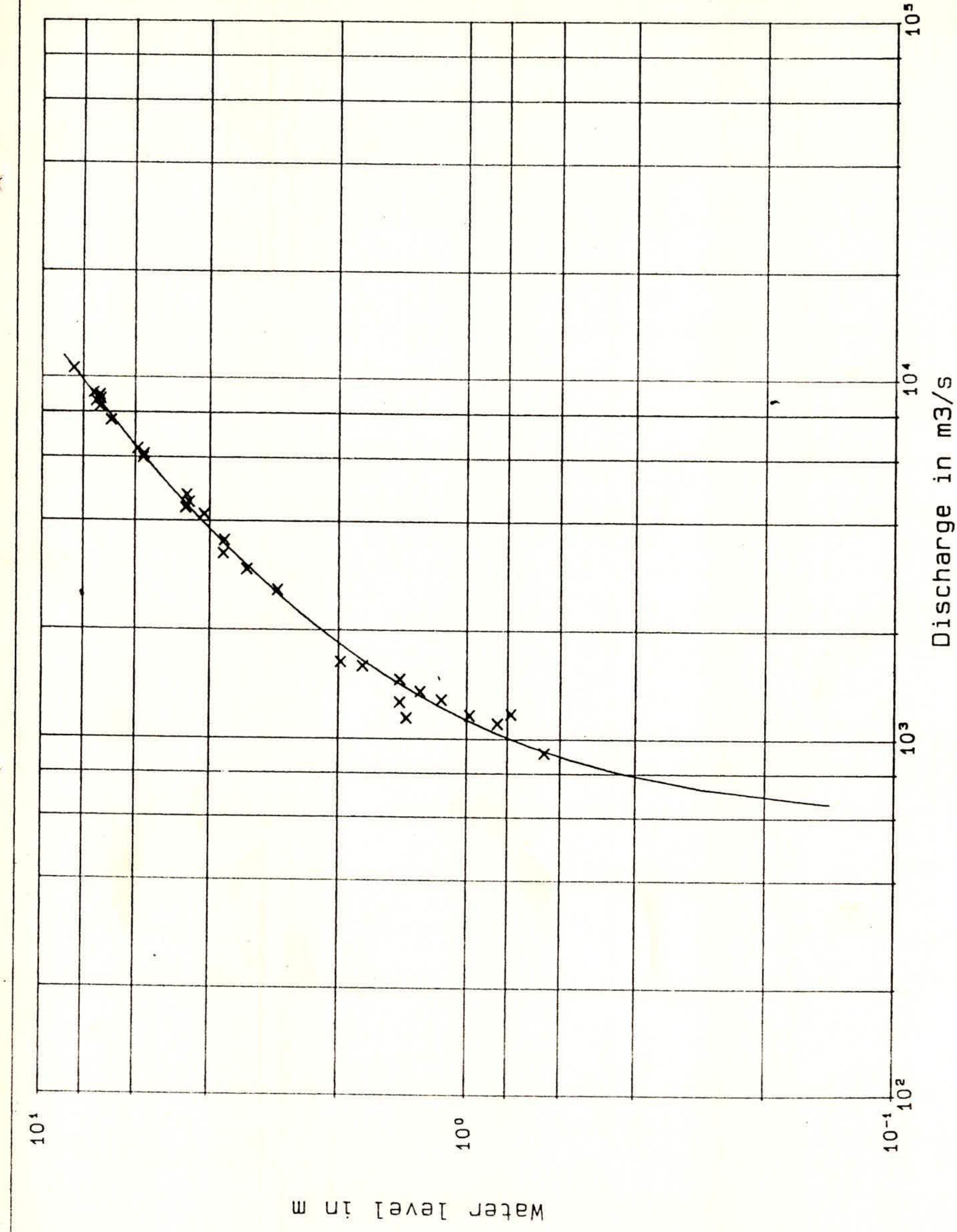


FIG.19 b Fitting of stage discharge data by a simple rating curve (Semi-log)

Fitting Simple Rating Curve GD Site at Haridwar (hypothetical)		Report
National Institute of Hydrology		FIG.

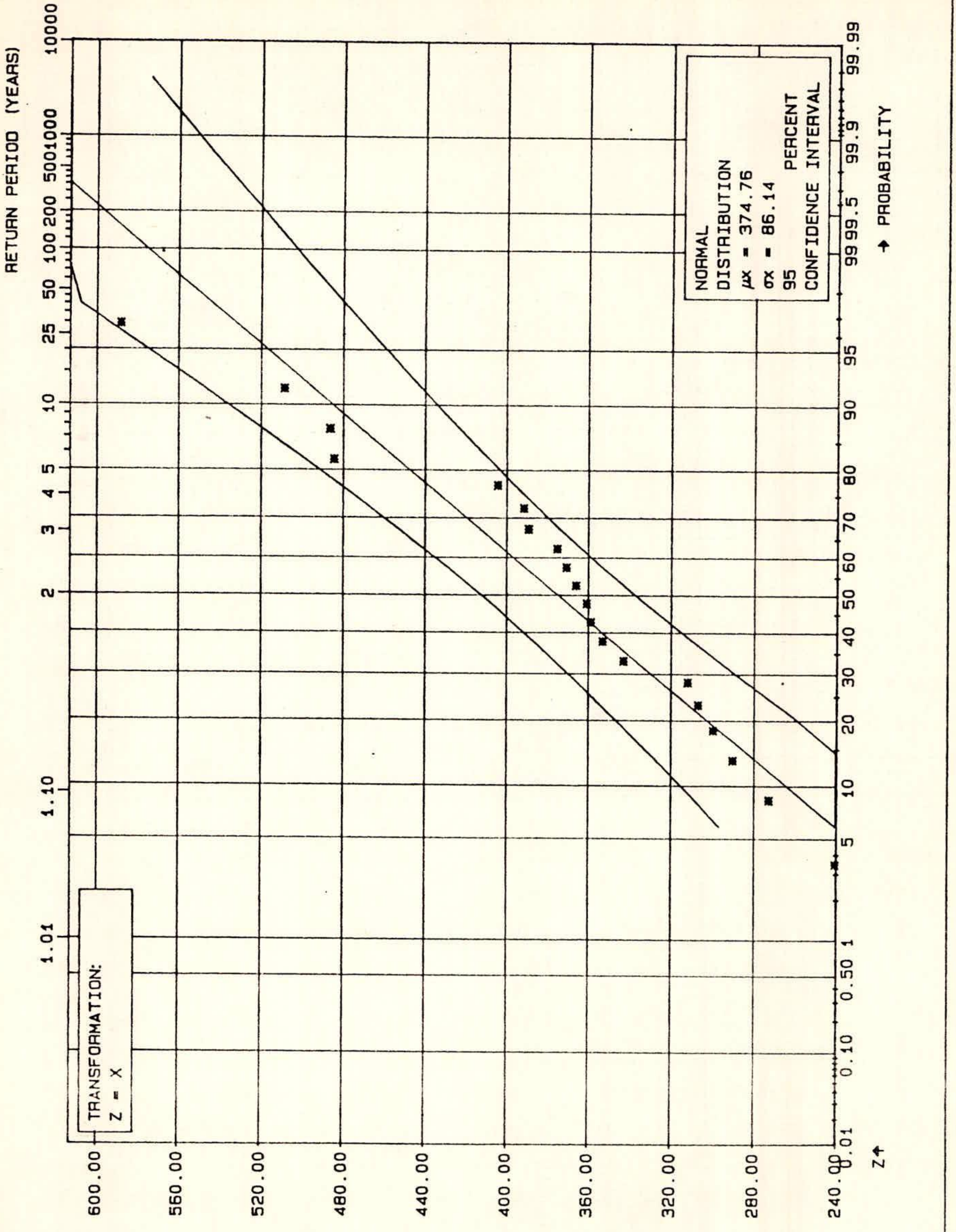
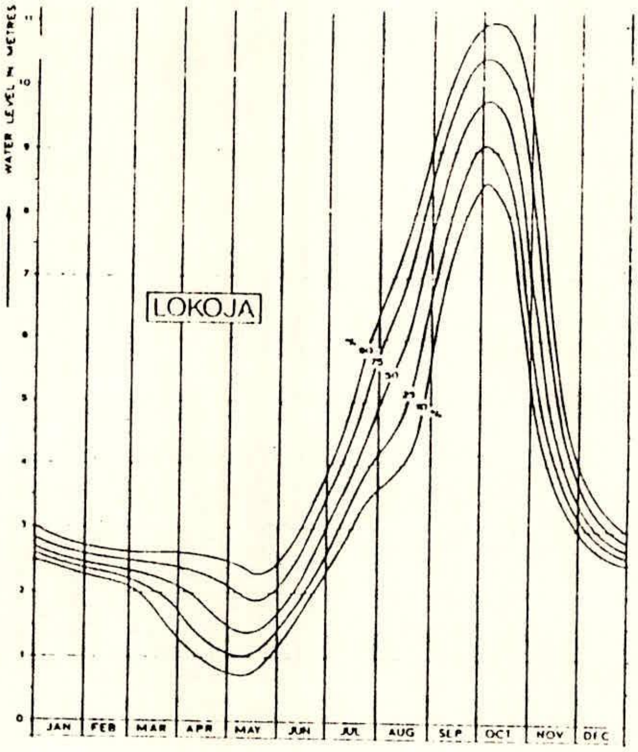
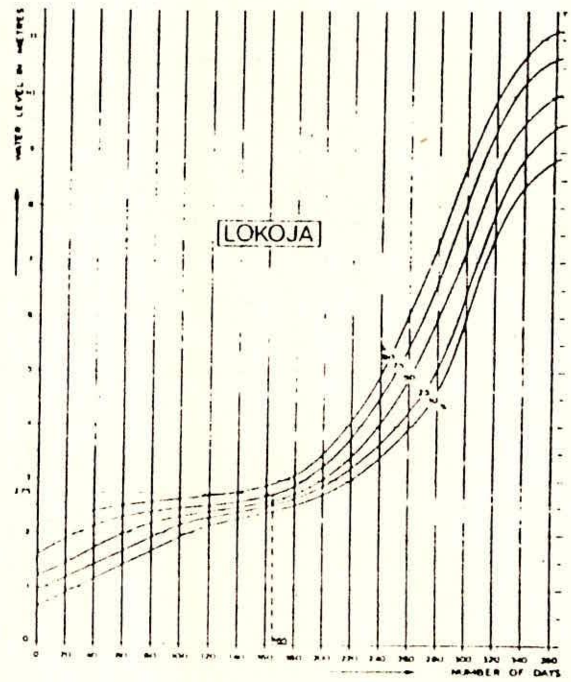


FIG.20 Example of plot of distribution function

Plot of Distribution Function		
	170201 PH2 1	
National Institute of Hydrology		FIG.

Frequency curves





Duration curves

Figure 21 Example of frequency and duration curves

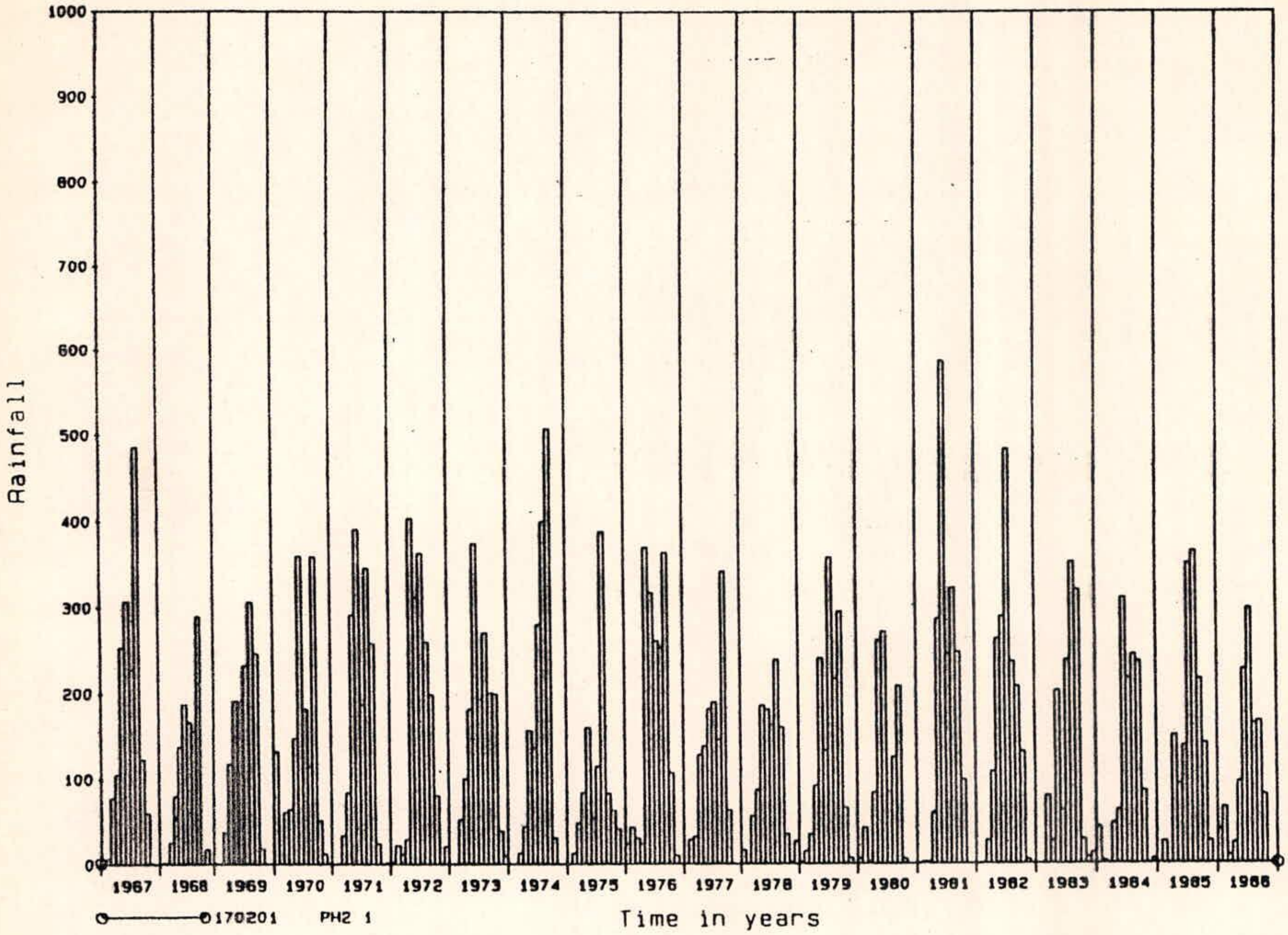


Figure 22 Time series of monthly rainfall

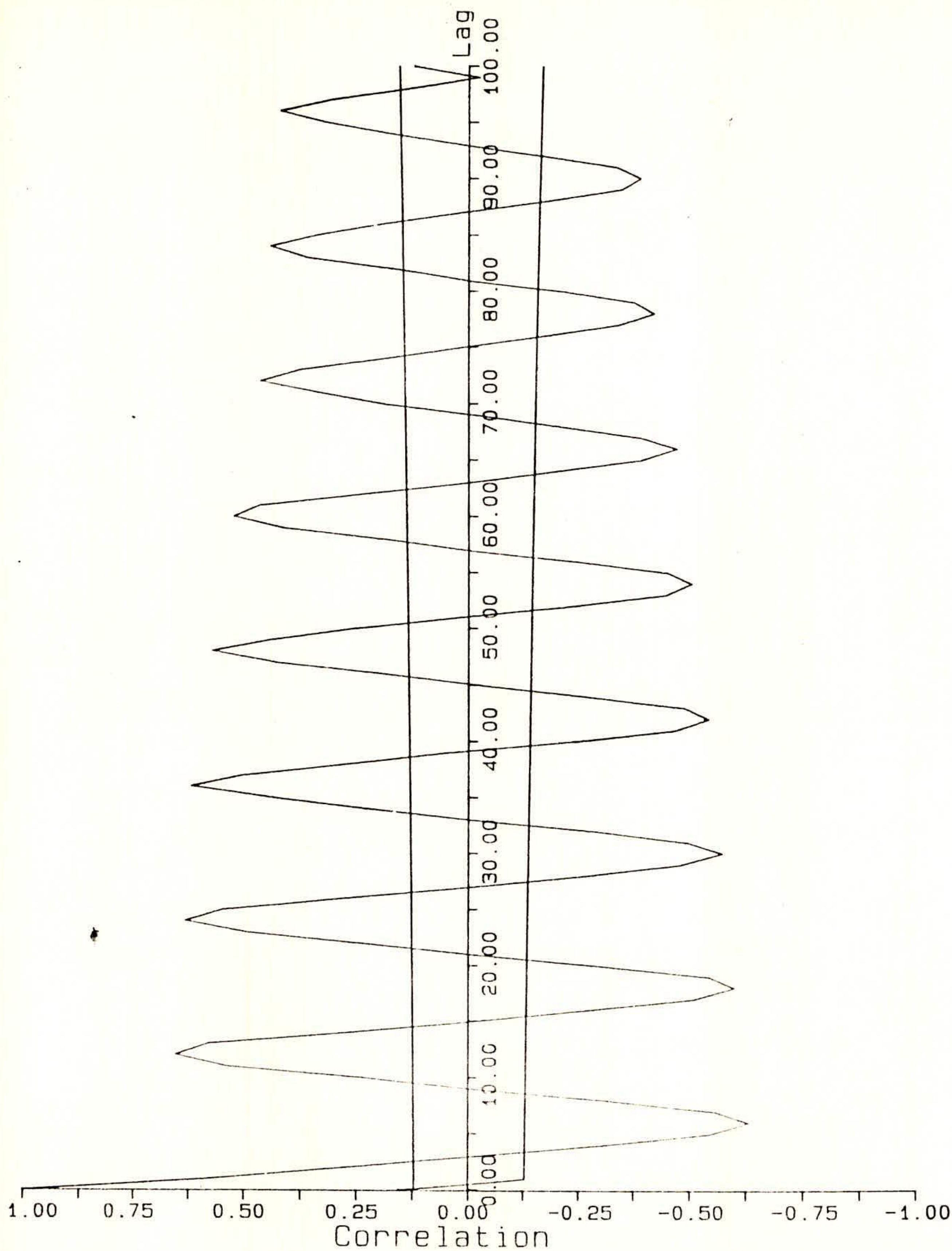


FIG. 23 Auto-Correlogram of monthly rainfall

Autocorrelogram
Series code 170201 PH2 1

National Institute of Hydrology

Report

FIG.

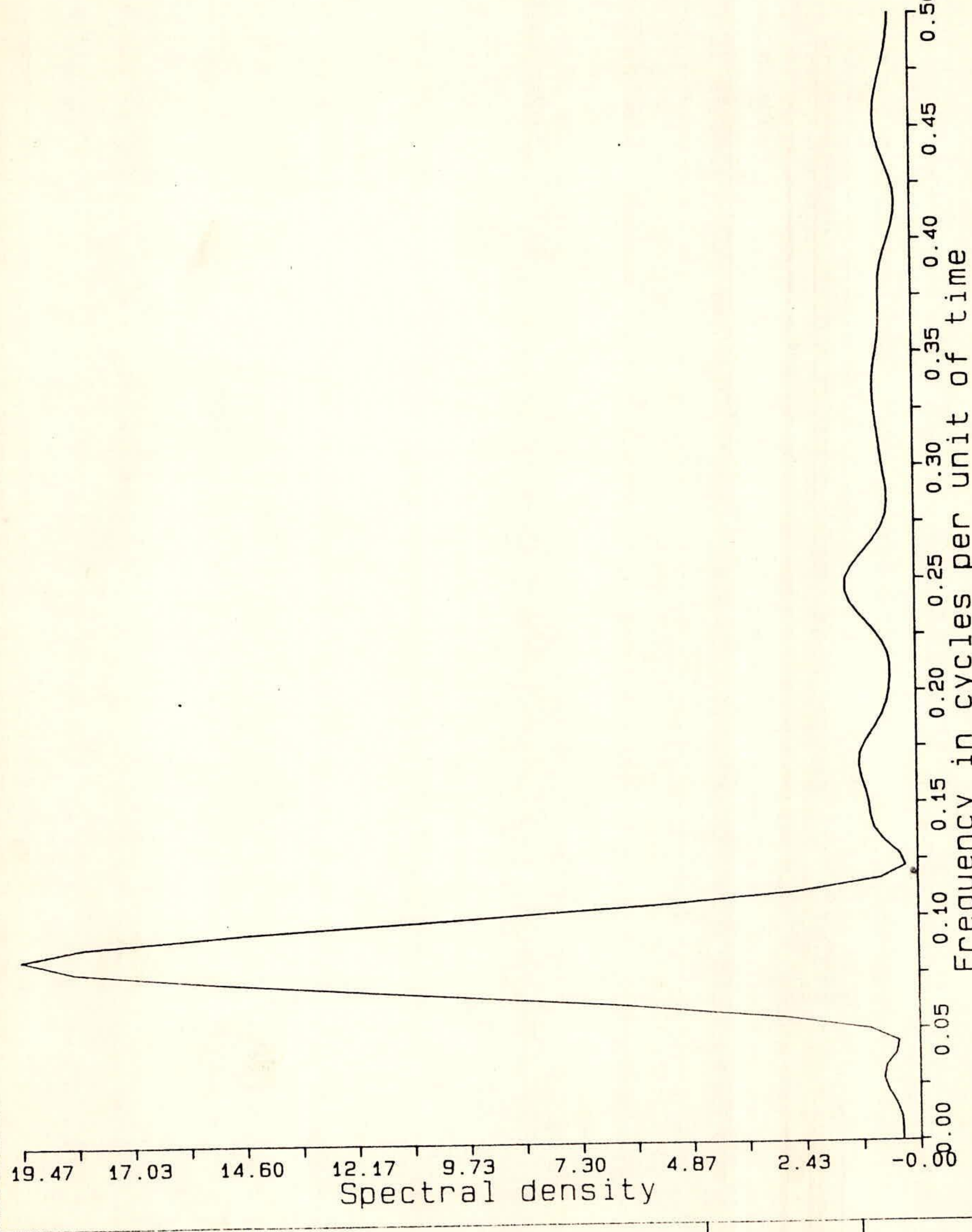


FIG. 24 Spectral density function of monthly rainfall

Spectral density function
Series code 170201 PH2 1

Report

National Institute of Hydrology

FIG.

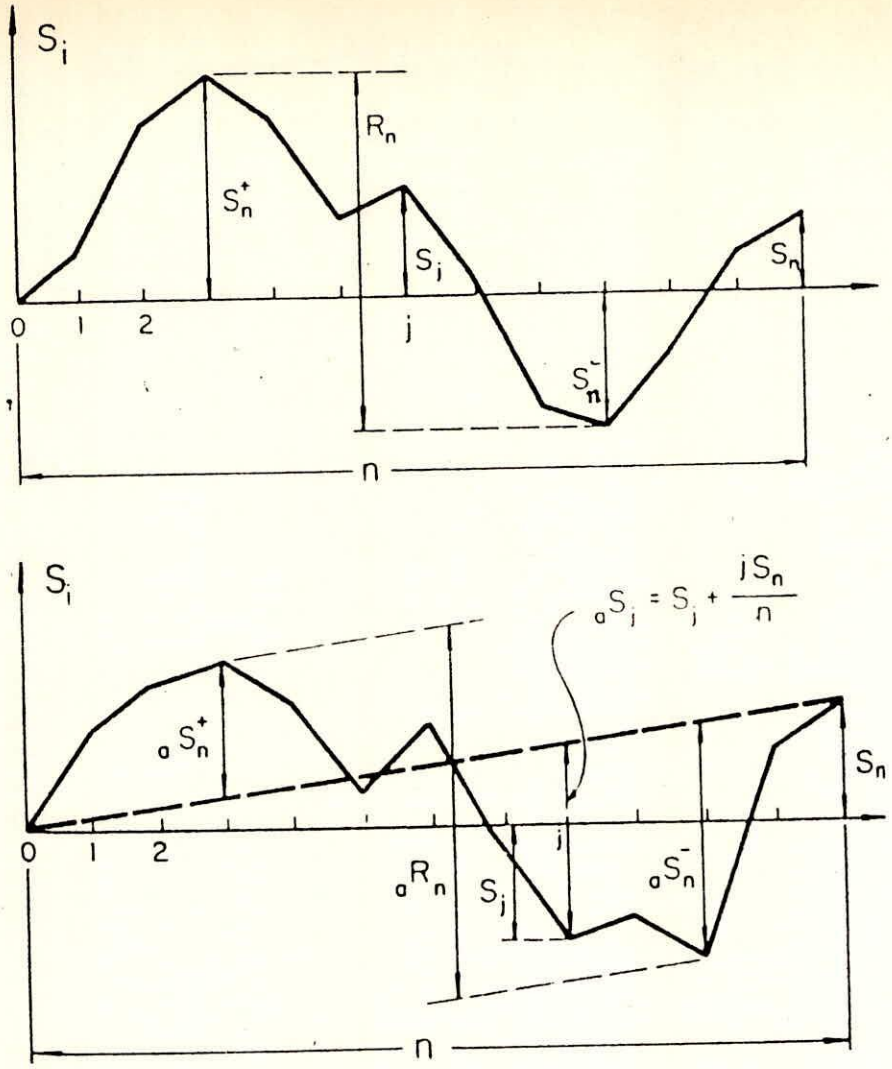


Figure 25 Definition sketch range quantities

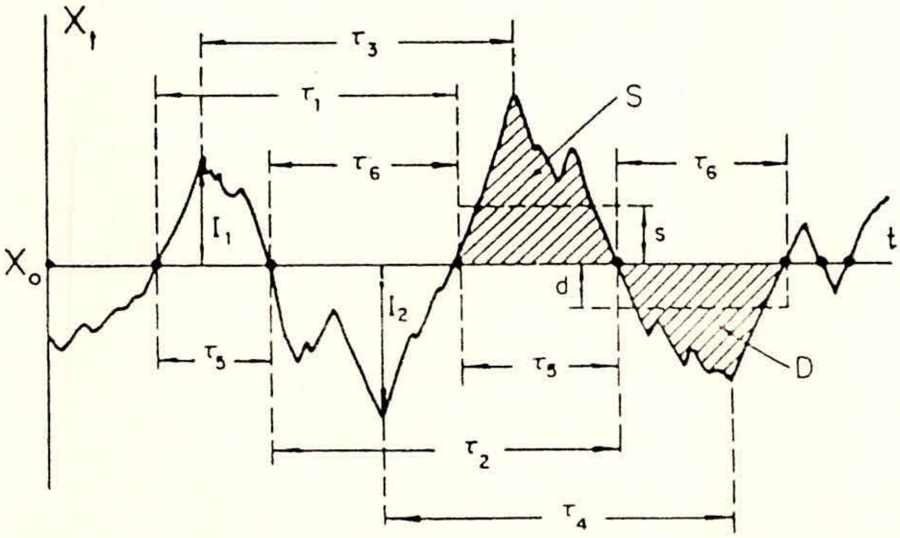


Figure 26 Definition sketch for run analysis

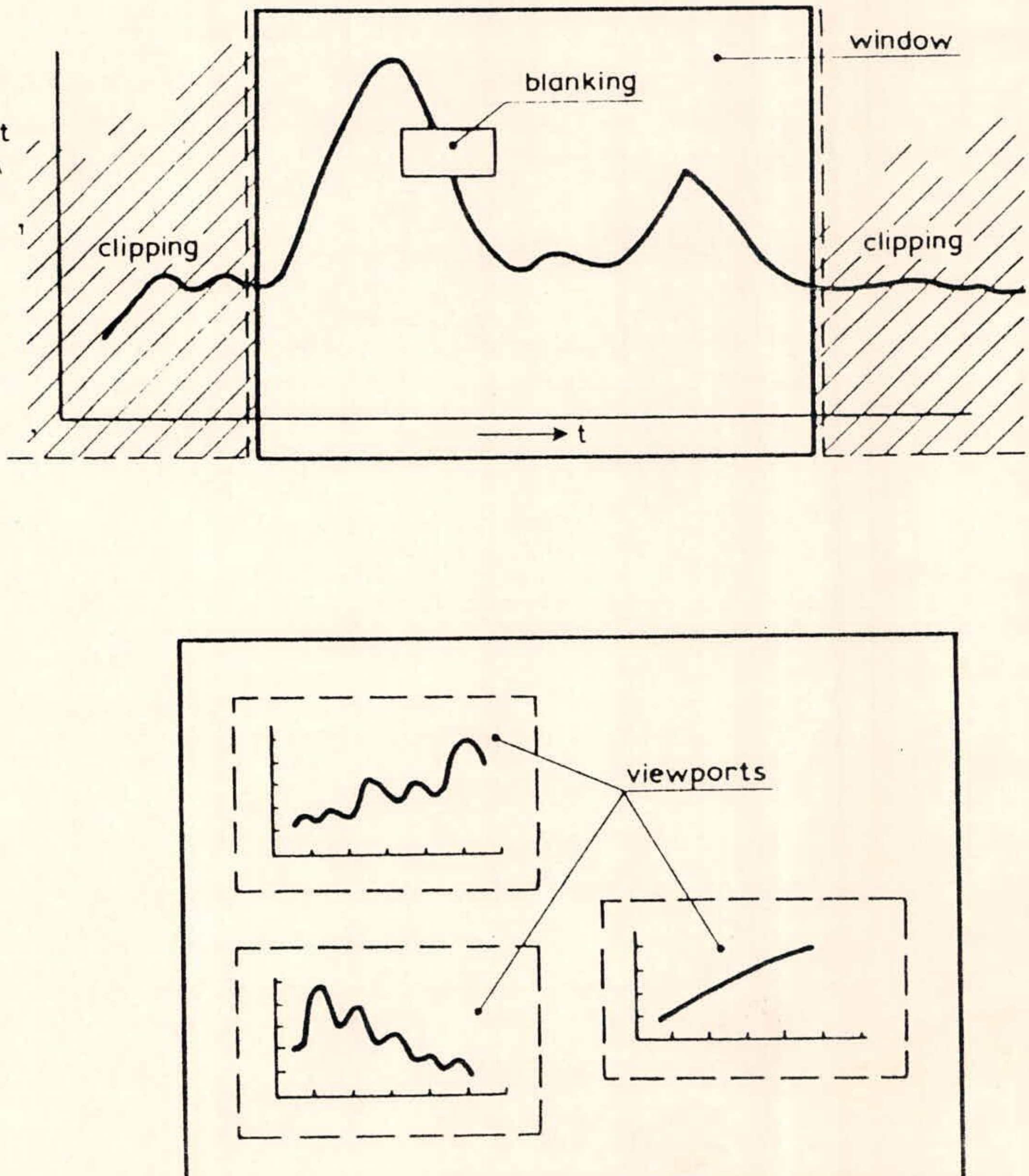


FIG. 27 Examples of graphical options

Monthly data of series 170201 PH2 1 Period 1967 - 1986

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
1967	.9	6.9	78.0	105.6	254.2	307.2	228.3	486.0	123.4	60.0	1.2	.0	1651.7
1968	2.3	1.7	25.9	80.3	138.0	187.8	167.3	157.0	290.0	.0	17.9	.0	1068.2
1969	.7	.0	37.6	118.1	191.3	190.6	233.2	306.6	247.3	18.5	.1	.0	1344.0
1970	132.8	.0	61.1	64.6	148.3	360.4	182.5	115.3	359.8	51.8	12.7	.0	1489.3
1971	.0	.0	33.9	84.5	292.1	391.2	187.7	346.4	259.4	24.6	.0	2.3	1622.1
1972	2.4	22.6	11.7	29.9	404.2	312.7	364.0	260.9	199.6	81.0	.7	21.3	1711.0
1973	.0	.0	52.8	101.1	182.2	374.7	193.5	271.6	201.6	200.2	39.6	9.8	1627.1
1974	.0	.0	13.6	44.4	157.8	137.4	280.7	400.4	508.3	31.1	.0	.0	1573.7
1975	.0	13.0	49.1	84.2	161.3	54.5	115.6	388.8	83.8	63.7	41.5	.0	1055.5
1976	24.5	43.7	31.2	24.0	370.2	317.7	262.3	254.8	364.2	108.3	10.8	.0	1811.7
1977	.0	29.3	33.4	129.4	139.5	182.7	190.7	147.1	342.4	64.0	.0	.0	1258.5
1978	17.0	.0	56.9	88.0	186.9	182.2	164.8	240.4	161.6	35.7	2.3	27.6	1163.4
1979	2.5	15.8	35.2	92.8	242.0	134.2	358.4	218.0	295.9	66.8	7.5	.0	1469.1
1980	6.5	43.2	3.0	84.2	262.6	272.5	85.0	126.5	209.2	6.2	.0	.0	1098.9
1981	.0	3.0	3.0	60.9	287.4	587.8	246.2	322.7	248.9	98.9	.0	.0	1858.8
1982	.0	.0	27.8	108.8	264.2	290.2	484.1	238.1	208.9	132.6	5.3	.0	1760.0
1983	.0	1.2	80.4	27.4	203.6	63.5	239.6	352.6	321.0	29.5	7.7	13.5	1340.0
1984	43.9	4.0	.0	48.8	63.8	311.7	217.7	246.3	238.4	86.4	.0	6.5	1267.5
1985	.0	26.8	.0	151.5	93.6	139.7	350.9	365.5	217.2	142.4	27.3	.0	1514.9
1986	40.3	66.1	10.0	25.2	96.3	228.6	299.4	165.3	168.1	81.7	.0	.0	1181.0
Effective	20	20	20	20	20	20	20	20	20	20	20	20	20
Missing	0	0	0	0	0	0	0	0	0	0	0	0	0
Mean	13.7	13.9	32.2	77.7	207.0	251.4	242.6	270.5	252.4	69.2	8.7	4.1	1443.3
St.dev	30.3	18.6	24.2	35.6	87.5	124.0	91.8	98.1	94.0	48.7	12.8	7.8	250.7
Stdv/mean	2.21	1.34	.75	.46	.42	.49	.38	.36	.37	.70	1.46	1.92	.17
Minimum	.0	.0	.0	24.0	63.8	54.5	85.0	115.3	83.8	.0	.0	.0	1055.5
Maximum	132.8	66.1	80.4	151.5	404.2	587.8	484.1	486.0	508.3	200.2	41.5	27.6	1858.8

Completed data marked with *
Corrected data marked with +

Table 1 : Example of dedicated table.

Data of series 170201 PH3 1 Date from 1986 1 1 0 1 (period 1)

.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	6.8	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	8.3	.0
.0	.0	.0	.0	35.0	.0	7.3	5.1	44.6	33.6
.0	.0	.0	5.3	.0	.0	.0	.0	.0	8.5
20.7	.5	19.5	43.3	52.3	.0	.0	.0	.0	.0
.0	.0	11.5	.0	5.1	11.0	45.0	25.0	.0	.0
.0	3.9	75.0	7.2	3.5	.0	1.9	7.5	50.9	.0
7.7	.0	.0	60.3	4.5	1.9	2.1	.0	.0	.0
.0	.0	.0	.0	17.9	.0	6.9	4.9	1.2	15.3
.0	.0	.0	.0	.0	.0	.0	33.1	2.8	.0
56.2	33.8	2.5	12.0	2.3	.0	9.8	.0	.0	.0
.0	.0	5.0	.0	18.0	.0	14.7	.0	9.5	3.2
69.0	.0	.0	.0	.0	.0	3.6	.0	.0	.0
.0	.0	.0	8.2	50.0	.0	69.1	20.9	.0	.0
.0	.0	7.5	.0	.0	.0	.0	.0	31.2	28.5
12.5	.0	36.0	29.0	.0	.0	.0	.0	45.0	2.5
99.3	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	5.2	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	2.5	.0	.0	.0
41.5	.0	.0	.0	7.5	24.5	.0	.0	.0	15.8
17.2	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
16.5	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0					

Maximum = 99.300, minimum= .000, mean= 4.128

Completed data marked with *
Corrected data marked with +

Missing data signified by -1.

Table 2 : Example of data block.

Daily data and statistics of series 170201 PH3 1 Year = 1986

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	.0	.0	.0	.0	.0	.0	.0	5.0	.0	.0	.0	.0
2	.0	.0	.0	.0	.0	11.5	.0	.0	.0	.0	.0	.0
3	.0	.0	.0	.0	.0	.0	.0	18.0	.0	.0	.0	.0
4	.0	.0	.0	6.8	.0	5.1	17.9	.0	.0	.0	.0	.0
5	.0	.0	.0	.0	35.0+	11.0	.0	14.7	.0	.0	.0	.0
6	.0	.0	.0	.0	.0-	45.0+	6.9	.0	31.2+	.0	.0	.0
7	.0	.0	.0	.0	7.3	25.0	4.9	9.5	28.5	.0	.0	16.5
8	.0	.0	.0	.0	5.1	.0	1.2	3.2	12.5	.0	.0	.0
9	.0	.0	.0	.0	44.6+	.0	15.3	69.0*	.0	.0	.0	.0
10	.0	.0	.0	.0	33.6	.0	.0	.0-	36.0+	.0	.0	.0
11	.0	.0	.0	.0	.0-	3.9	.0	.0	29.0	.0	.0	.0
12	.0	.0	.0	.0	.0	75.0*	.0	.0	.0	.0	.0	.0
13	.0	.0	.0	.0	.0	7.2-	.0	.0	.0	.0	.0	.0
14	.0	.0	.0	.0	5.3	3.5	.0	.0	.0	2.5	.0	.0
15	.0	.0	.0	.0	.0	.0	.0	3.6	.0	.0	.0	.0
16	.0	.0	.0	.0	.0	1.9	.0	.0	45.0+	.0	.0	.0
17	.0	.0	.0	.0	.0	7.5	33.1+	.0	2.5-	.0	.0	.0
18	.0	.0	.0	.0	.0	50.9*	2.8-	.0	99.3*	41.5+	.0	.0
19	.0	.0	.0	.0	.0	.0-	.0	.0	.0-	.0-	.0	.0
20	.0	.0	.0	.0	8.5	7.7	56.2*	.0	.0	.0	.0	.0
21	.0	.0	.0	.0	20.7	.0	33.8	.0	.0	.0	.0	.0
22	.0	.0	.0	.0	.5	.0	2.5-	8.2	.0	7.5	.0	.0
23	.0	.0	.0	.0	19.5	60.3*	12.0	50.0*	.0	24.5	.0	.0
24	.0	.0	.0	.0	43.3	4.5-	2.3	.0-	.0	.0	.0	.0
25	.0	.0	.0	.0	52.3*	1.9	.0	69.1*	.0	.0	.0	.0
26	.0	.0	.0	.0	.0-	2.1	9.8	20.9-	.0	.0	.0	.0
27	.0	.0	.0	.0	.0	.0	.0	.0	.0	15.8	.0	.0
28	.0	.0	.0	.0	.0	.0	.0	.0	.0	17.2	.0	.0
29	.0	*********	.0	8.3	.0	.0	.0	.0	.0	.0	.0	.0
30	.0	*********	.0	.0	.0	.0	.0	.0	5.2	.0	.0	.0
31	.0	*********	.0	*********	.0	*********	.0	7.5	*********	.0	*********	.0
Data	31	28	31	30	31	30	31	31	30	31	30	31
Eff.	31	28	31	30	31	30	31	31	30	31	30	31
Miss	0	0	0	0	0	0	0	0	0	0	0	0
Sum	.00	.00	.00	15.10	275.70	324.00	198.70	278.70	289.20	109.00	.00	16.50
Mean	.00	.00	.00	.50	8.89	10.80	6.41	8.99	9.64	3.52	.00	.53
Min.	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Max.	.00	.00	.00	8.30	52.30	75.00	56.20	69.10	99.30	41.50	.00	16.50
High	45.00	45.00	45.00	45.00	45.00	45.00	45.00	45.00	45.00	45.00	45.00	45.00
Numb	0	0	0	0	1	3	1	3	1	0	0	0
Low	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Numb	0	0	0	0	0	0	0	0	0	0	0	0

Annual values:

Data	365	* Sum	1506.900	* Minimum	.000	* Too low	0
Effective	365	* Mean	4.128	* Maximum	99.300	* Too high	9
Missing	0						

Exceedance of:

- Lower bound (.00) marked with *
- Upper bound (45.00) marked with *
- Rate of rise (30.00) marked with +
- Rate of fall (30.00) marked with -

Table 3 : Example of screening of daily data.

Tabulation of series, Year 1986

Year	mth	day	hr	si	170201 PH	170202 PH	170203 PH	170205 PH	170206 PH	170209 PH
					==========Data==========					
1986	6	1	0	1	.0	.0	.0	.0	.0	.0
1986	6	2	0	1	11.5	11.5	.0	7.2	3.4	.0
1986	6	3	0	1	.0	.0	.0	.0	6.8	.0
1986	6	4	0	1	5.1	5.1	.0	.0	3.6	.0
1986	6	5	0	1	11.0	11.0	18.5	11.3	29.8	2.0
1986	6	6	0	1	45.0	45.0	.0	8.5	63.1	3.2
1986	6	7	0	1	25.0	25.0	72.3	13.8	18.6	.0
1986	6	8	0	1	.0	.0	.0	.0	5.8	.0
1986	6	9	0	1	.0	.0	14.5	.0	.8	.0
1986	6	10	0	1	.0	.0	12.0	5.1	.4	6.0
1986	6	11	0	1	3.9	3.9	8.5	.0	3.3	.0
1986	6	12	0	1	75.0	75.0	143.5	95.3	102.6	30.0
1986	6	13	0	1	7.2	7.2	21.5	7.8	15.2	31.8
1986	6	14	0	1	3.5	3.5	.0	10.2	1.6	18.0
1986	6	15	0	1	.0	.0	.0	.0	.0	.0
1986	6	16	0	1	1.9	1.9	.0	.0	.1	.0
1986	6	17	0	1	7.5	7.5	17.8	3.5	14.0	.0
1986	6	18	0	1	50.9	50.9	30.0	22.5	37.1	64.5
1986	6	19	0	1	.0	.0	.0	1.5	1.4	11.2
1986	6	20	0	1	7.7	7.7	.0	.0	2.3	.0
1986	6	21	0	1	.0	.0	.0	.0	.0	.0
1986	6	22	0	1	.0	.0	.0	.0	.0	.0
1986	6	23	0	1	60.3	60.3	85.4	5.3	58.5	16.5
1986	6	24	0	1	4.5	4.5	11.2	.0	5.0	5.0
1986	6	25	0	1	1.9	1.9	91.5	.0	2.6	.0
1986	6	26	0	1	2.1	2.1	8.3	.0	7.9	.0
1986	6	27	0	1	.0	.0	.0	.0	.0	3.0
1986	6	28	0	1	.0	.0	.0	.0	.1	.0
1986	6	29	0	1	.0	.0	.0	.0	.0	.0
1986	6	30	0	1	.0	.0	.0	.0	.0	.0

Table 4 : Tabulation of daily rainfall data.

Double mass analysis of station 1 Base
Station 2 Test

1 Period	2 Amount mm	3 Station 1 Cum mm	4 Perc	5 Amount mm	6 Station 2 Cum mm	7 Perc	8 Ratios (3)/(6) -	9 (4)/(7) -
1967	1562.	1562.	5.13	1266.	1266.	2.84	1.23	1.80
1968	1399.	2962.	9.73	1500.	2766.	6.22	1.07	1.56
1969	1506.	4468.	14.67	1546.	4312.	9.69	1.04	1.51
1970	2030.	6498.	21.34	2508.	6820.	15.32	.95	1.39
1971	1583.	8081.	26.54	1809.	8629.	19.39	.94	1.37
1972	1423.	9503.	31.21	1823.	10452.	23.48	.91	1.33
1973	1455.	10959.	35.99	1757.	12208.	27.43	.90	1.31
1974	1465.	12423.	40.80	1815.	14023.	31.51	.89	1.29
1975	1734.	14157.	46.49	2324.	16347.	36.73	.87	1.27
1976	1591.	15748.	51.72	1399.	17747.	39.87	.89	1.30
1977	1148.	16896.	55.49	1583.	19330.	43.43	.87	1.28
1978	1688.	18584.	61.03	2308.	21638.	48.61	.86	1.26
1979	1228.	19812.	65.06	1342.	22980.	51.63	.86	1.26
1980	1740.	21552.	70.78	3451.	26431.	59.38	.82	1.19
1981	1718.	23270.	76.42	3297.	29727.	66.79	.78	1.14
1982	1575.	24845.	81.59	3006.	32734.	73.54	.76	1.11
1983	1282.	26126.	85.80	2855.	35589.	79.96	.73	1.07
1984	1456.	27582.	90.58	3240.	38829.	87.24	.71	1.04
1985	1270.	28852.	94.75	2643.	41472.	93.18	.70	1.02
1986	1598.	30450.	100.00	3037.	44509.	100.00	.68	1.00

Total number of periods analysis: 20

Table 5 : Example of result of double mass analysis.

Program HYCUR

Processing of discharge measurements

Station : VIEN

Number : 2
Date :1988 7 28, time: 0 0

Computation of total flow by mean-section method

Distance ref.point-left bank (m)	:	4.00
Width of river (m)	:	26.55
Level of gauge zero (m)	:	.00
Water level begin (m)	:	1.70
Water level end (m)	:	1.80
Dh/Dt (m/day)	:	2.40

Current meter type : OTT12345

Rating of current meter:

<=rev	a	b
.52	.2521	.0220
3.62	.2644	.0127
-11.08	.2728	.0022

Verticals

Vert. no.	Dist. ref.p m	Depth m	Mean vel. m/s	Disch. unit-w m3/s/m
0	4.00	.00	.000	.000
1	5.00	.31	.189	.059
2	6.00	.40	.214	.086
3	7.00	.51	.234	.119
4	8.00	.85	.238	.202
5	9.00	1.23	.231	.284
6	10.00	1.58	.236	.373
7	11.00	1.69	.247	.417
8	12.00	1.71	.258	.442
9	13.00	1.87	.282	.527
10	14.00	1.84	.282	.518
11	15.00	1.71	.273	.467
12	16.00	1.65	.258	.426
13	17.00	1.50	.253	.379
14	18.00	1.36	.237	.322
15	19.00	1.19	.220	.262
16	20.00	1.17	.207	.243
17	21.00	.92	.212	.195
18	22.00	.81	.185	.150
19	23.00	.70	.182	.127
20	24.00	.63	.164	.103
21	25.00	.55	.150	.083
22	26.00	.48	.124	.060
23	26.50	.40	.101	.040
0	30.55	.00	.000	.000

Sections

Section no.	Width m	Area m2	Discharge m3/s
1	1.00	.16	.01[illegible]
2	1.00	.36	.07[illegible]
3	1.00	.45	.102
4	1.00	.68	.160
5	1.00	1.04	.244
6	1.00	1.40	.328
7	1.00	1.63	.395
8	1.00	1.70	.429
9	1.00	1.79	.48[illegible]
10	1.00	1.86	.52[illegible]
11	1.00	1.78	.49[illegible]
12	1.00	1.68	.4[illegible]
13	1.00	1.58	.4[illegible]
14	1.00	1.43	.3[illegible]
15	1.00	1.28	.2[illegible]
16	1.00	1.18	.25[illegible]
17	1.00	1.04	.21[illegible]
18	1.00	.87	.172
19	1.00	.75	.138
20	1.00	.66	.115
21	1.00	.59	.093
22	1.00	.51	.071
23	.50	.22	.025
24	4.05	.81	.04[illegible]
Totals:	26.55	25.45	5.85[illegible]

Wetted perimeter (m) : 26.93
Average velocity (m/s): .23

Hydraulic radius (m): .94

Corrected water level (m) : 1.74

Table 6 : Discharge computation by meansection method from point velocity

Analysis of stage-discharge data

Station name : VIEN

Data from 1986 1 1 to 1986 12 31

Given boundaries for computation of rating curve(s)

interval	lower bound	upper bound	nr. of data
1	.000	9.000	29

Power type of equation q=c*(h+a)**b is used

Boundaries / coefficients

lower bound	upper bound	a	b	c
.00	9.00	2.625	2.012	.8464E+02

Gauge Zero on 1986 2 3 = .000 m

Number of data = 29
h minimum = .65 meas.nr = 6
h maximum = 8.39 meas.nr = 18
q minimum = 910.000 meas.nr = 6
q maximum = 10600.000 meas.nr = 18

Number	W level M	Q meas M3/S	Q comp M3/S	Diff M3/S	Rel.diff 0/0	Semr 0/0
1	1.430	1260.000	1415.930	-155.930	-11.01	3.58
2	1.430	1460.000	1415.930	44.070	3.11	3.58
3	1.140	1280.000	1219.561	60.439	4.96	3.93
4	1.280	1350.000	1312.519	37.481	2.86	3.76
5	.840	1100.000	1031.924	68.076	6.60	4.35
6	.650	910.000	921.236	-11.236	-1.22	4.64
7	.980	1160.000	1117.527	42.473	3.80	4.14
8	1.380	1140.000	1381.022	-241.022	-17.45	3.64
9	.780	1170.000	996.288	173.712	17.44	4.44
10	1.750	1590.000	1649.729	-59.729	-3.62	3.26
11	1.970	1630.000	1820.890	-190.890	-10.48	3.07
12	3.260	2920.000	2995.716	-75.716	-2.53	2.53
13	3.700	3230.000	3463.433	-233.433	-6.74	2.53
14	4.450	4500.000	4339.370	160.630	3.70	2.70
15	5.900	6320.000	6314.552	5.448	.09	3.31
16	5.690	6110.000	6005.471	104.529	1.74	3.21
17	7.280	8270.000	8539.892	-269.892	-3.16	4.00
18	8.390	10600.000	10574.790	25.206	.24	4.54
19	7.420	8560.001	8784.502	-224.501	-2.56	4.07
20	7.230	8690.000	8453.372	236.628	2.80	3.97
21	7.520	9040.001	8961.353	78.648	.88	4.12
22	7.260	8920.000	8505.230	414.770	4.88	3.99
23	6.830	7570.000	7777.170	-207.170	-2.66	3.77
24	5.710	5970.000	6034.571	-64.571	-1.07	3.22
25	4.500	4710.000	4401.296	308.704	7.01	2.72
26	4.100	4160.000	3918.249	241.751	6.17	2.60
27	3.670	3520.000	3430.459	89.541	2.61	2.53
28	4.540	4340.000	4451.154	-111.154	-2.50	2.73
29	2.760	2560.000	2505.613	54.387	2.17	2.63

Overall standard error = 6.623

Statistics per interval

Interval	Lower bound	Upper bound	Nr.of data	Standard error
1	.000	9.000	29	6.63

Table 7 : Summary of computational results.

Series code = 170201 PH2 1

First year = 1967

Last year = 1986

Annual maximum values are analysed

Series code = 170201 PH2 1

Year	Maximum
1967	4.860000E+02
1968	2.900000E+02
1969	3.066000E+02
1970	3.604000E+02
1971	3.912000E+02
1972	4.042000E+02
1973	3.747000E+02
1974	5.083000E+02
1975	3.888000E+02
1976	3.702000E+02
1977	3.424000E+02
1978	2.404000E+02
1979	3.584000E+02
1980	2.725000E+02
1981	5.878000E+02
1982	4.841000E+02
1983	3.526000E+02
1984	3.117000E+02
1985	3.655000E+02
1986	2.994000E+02

Values for distinct return periods

Return per.	prob(xi<x) p	value x	st. dev. x	95% confidence interval	
				lower	upper
2	.50000	374.760	19.263	.331	.669
5	.80000	447.247	22.414	.630	.912
10	.90000	485.174	25.997	.755	.969
25	.96000	525.606	30.658	.854	.993
50	.98000	551.717	33.969	.900	.998
100	.99000	575.200	37.087	.931	.999
250	.99600	603.255	40.943	.957	1.000
500	.99800	622.728	43.683	.970	1.000
1000	.99900	640.992	46.293	.979	1.000
2500	.99960	663.603	49.566	.987	1.000
5000	.99980	679.736	51.926	.991	1.000
10000	.99990	695.139	54.195	.994	1.000

Fitting the normal distribution function

Series code 170201 PH2 1

Number of data	=	20
Mean	=	374.760
Standard deviation	=	86.145
Skewness	=	.880
Kurtosis	=	3.665

Table 8 : Example output of analysis of annual maximum monthly rainfall data.

Compilation of frequency and duration curves
==

Series code 170201 PH2 1
Start date 1967 1 0 0 1 End date 1986 12 0 0 1
No. of values in period 12

Frequencies: .10 .25 .50 .75 .90

Frequency curves 170201 PH2 1

Element	Nr.data	Frequency .10	.25	.50	.75	.90
1	20	.00	.00	.80	14.37	43.54
2	20	.00	.00	3.50	25.75	43.65
3	20	.30	10.43	32.30	51.87	76.31
4	20	25.42	45.50	84.20	104.47	128.27
5	20	93.87	141.70	189.10	263.80	362.39
6	20	70.57	150.32	250.55	316.45	389.55
7	20	120.52	183.80	230.75	294.73	363.44
8	20	128.56	178.48	257.85	351.05	399.24
9	20	127.22	200.10	242.85	314.72	363.76
10	20	7.43	29.90	63.85	95.77	141.42
11	20	.00	.00	1.75	12.22	38.37
12	20	.00	.00	.00	5.45	20.52

Table 9 Example of frequency curves

Time series analysis
====================

Autocovariance and autocorrelation analysis
===

Series = 170201   PH2 1
Date of first element = 1967 1 0 0 1
Date of last element = 1986 12 0 0 1

COV = autocovariance function
COR = autocorrelation function
CLP = upper conf. limit zero correlation (95 %)
CLN = lower conf. limit zero correlation (95 %)

LAG	COV	COR	CLP	CLN
0	.1635E+05	1.0000	.1213	.1562
1	.9375E+04	.5735	.1216	-.1299
2	.4545E+04	.2780	.1218	-.1302
3	-.1355E+03	-.0083	.1220	-.1304
4	-.5034E+04	-.3079	.1223	-.1307
5	-.9029E+04	-.5523	.1225	-.1310
6	-.1038E+05	-.6349	.1228	-.1313
7	-.9127E+04	-.5583	.1230	-.1316
8	-.5174E+04	-.3165	.1233	-.1319
9	-.3261E+03	-.0199	.1235	-.1322
10	.3831E+04	.2343	.1238	-.1324
11	.8801E+04	.5383	.1240	-.1327
12	.1067E+05	.6524	.1243	-.1330
13	.9383E+04	.5740	.1246	-.1333
14	.4337E+04	.2653	.1248	-.1336
15	-.1568E+03	-.0096	.1251	-.1339
16	-.4435E+04	-.2713	.1253	-.1342
17	-.8397E+04	-.5137	.1256	-.1345
18	-.9853E+04	-.6027	.1259	-.1349
19	-.8894E+04	-.5441	.1262	-.1352
20	-.4746E+04	-.2903	.1264	-.1355
21	-.2919E+03	-.0179	.1267	-.1358
22	.3811E+04	.2331	.1270	-.1361
23	.8091E+04	.4950	.1272	-.1364
24	.1031E+05	.6310	.1275	-.1367
25	.8859E+04	.5420	.1278	-.1371
26	.4638E+04	.2837	.1281	-.1374
27	.5131E+02	.0031	.1284	-.1377
28	-.4309E+04	-.2636	.1287	-.1380
29	-.7889E+04	-.4826	.1289	-.1384
30	-.9409E+04	-.5756	.1292	-.1387
31	-.8106E+04	-.4959	.1295	-.1390
32	-.4778E+04	-.2923	.1298	-.1394
33	-.3567E+03	-.0218	.1301	-.1397
34	.3528E+04	.2158	.1304	-.1401
35	.7075E+04	.4328	.1307	-.1404
36	.1011E+05	.6183	.1310	-.1408
37	.8106E+04	.4959	.1313	-.1411

Table 10 : Example of output : auto-correlogram of monthly rainfall

Time series analysis
====================

Spectral analysis
=================

Series =170201 PH2 1
Date of first element = 1967 1 0 0 1
Date of last element = 1986 12 0 0 1

Truncation lag = 30
Number of frequency points= 90

Bandwith = .0444
Degr.frdom = 21

ASPEC = variance spectrum
LOG SPEC = logaritm of ASPEC
DSPEC = spectral density

NR	FREQUENCY	ASPEC	LOGSPEC	DSPEC
0	.0000	.6134E+04	3.7877	.3752
1	.0056	.6147E+04	3.7887	.3760
2	.0111	.6629E+04	3.8215	.4055
3	.0167	.8278E+04	3.9179	.5064
4	.0222	.1081E+05	4.0337	.6610
5	.0278	.1247E+05	4.0960	.7630
6	.0333	.1136E+05	4.0554	.6950
7	.0389	.7916E+04	3.8985	.4843
8	.0444	.6981E+04	3.8439	.4270
9	.0500	.1727E+05	4.2374	1.0567
10	.0556	.4765E+05	4.6780	2.9147
11	.0611	.1016E+06	5.0067	6.2123
12	.0667	.1728E+06	5.2376	10.5728
13	.0722	.2457E+06	5.3904	15.0296
14	.0778	.2998E+06	5.4768	18.3368
15	.0833	.3182E+06	5.5027	19.4656
16	.0889	.2950E+06	5.4699	18.0485
17	.0944	.2377E+06	5.3760	14.5402
18	.1000	.1637E+06	5.2141	10.0142
19	.1056	.9322E+05	4.9695	5.7024
20	.1111	.4113E+05	4.6141	2.5158

Table 11 : Example of output : Autospectrum of monthly rainfall

Station Characteristics of station : OBSERVATRY

Station name : NIH OBSERVATRY   Latitude : 00x 00' 00" North   Longitude : 000x 00' 00" East
River : GANGA   Altitude : .00 m   Catchment area : .000 km}
Province : HARIDWAR   Country : INDIA   Agency : N.I.H

NIH OBSERVATRY

RAINFALL HISTORICAL, YEAR 1987

DAY	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	.0	.0	.0	.0	18.0	.0	.0	.0	.0	.0	.0	.0
2	.0	.0	.0	.0	7.5	.0	.0	.0	.0	.0	.0	.0
3	.0	.0	.0	.0	.0	.0	.0	.0	1.4	.0	.0	.0
4	.0	.0	.0	.0	.0	.0	.0	3.4	10.8	.0	.0	.0
5	.0	.0	.0	.0	.0	.0	.0	.0	2.0	.0	.0	.0
6	.0	.0	.0	.0	28.2	.0	3.2	.0	.2	.0	.0	.0
7	.0	.0	.0	.0	9.2	.0	.0	.0	1.8	.0	.0	.0
8	.0	.0	.0	5.4	1.3	78.2	.0	1.0	.0	.0	.0	.0
9	.0	.0	.0	.0	29.3	30.2	.0	.0	.0	.0	.0	.0
10	.0	.0	.0	(-)	.0	2.4	.0	2.0	10.5	.0	.0	.0
11	.0	.0	2.2	(-)	3.0	.0	.0	.0	.0	.0	.0	4.6
12	.0	.0	.0	.0	7.6	.0	.0	14.0	.0	.0	.0	5.0
13	.0	.0	.2	.0	3.4	.0	1.8	9.2	.0	.0	.0	.2
14	.0	.0	.0	.0	.0	.0	.6	.0	.0	.0	.0	.0
15	.0	33.0	.0	.0	.0	.0	.2	.0	.0	.0	.0	.0
16	11.0	.7	.0	.0	.0	.0	.0	1.0	.0	.0	.0	.0
17	21.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
18	.1	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
19	.0	.0	2.2	.0	.0	.0	.0	36.6	.0	4.2	.0	.0
20	.0	.0	.0	.0	.0	.0	.2	.0	.0	.0	.0	.0
21	.0	.0	10.2	.0	.0	3.4	.0	.0	.0	.0	.0	.0
22	.0	.0	1.2	.0	4.2	.0	.0	.0	.0	.0	.0	.0
23	.0	.4	.0	.0	.2	.0	.0	.0	.0	.0	.0	.0
24	.0	11.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
25	.0	.0	.1	.0	.0	.0	.0	2.6	.0	.0	.0	.0
26	.0	.0	.0	.0	.0	.0	39.2	72.4	.0	.0	.0	.0
27	.0	.0	.0	.0	.0	.0	.0	1.1	.0	.0	.0	.0
28	.0	.0	.0	.0	.0	.0	.0	8.2	.0	.0	.0	.0
29	.0		.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
30	.0		.0	7.0	.0	.0	.0	22.8	.0	.0	.0	.0
31	.0		.0		.0		.0	1.4		.0		.0
Total	32.1	45.1	16.1	12.4	111.9	114.2	45.2	175.7	26.7	4.2	.0	9.8
Mean	1.0	1.6	.5	.4	3.6	3.8	1.5	5.7	.9	.1	.0	.3
Max.	21.0	33.0	10.2	7.0	29.3	78.2	39.2	72.4	10.8	4.2	.0	5.0
Min.	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0

Sum: 593.4   Mean: 1.6   Maximum: 78.2   Minimum: .0

Legend:
(-) Signifies missing data
T Signifies traces

Table 12 : Example of report.